The Dissertation Committee for Jake Allen LeVieux certifies that this is the approved version of the following dissertation:

Structure-Based Mechanism of Hydratase-Aldolases in the Type I Aldolase Superfamily and Structures of Tautomerase Superfamily Members

Committee:

Christian P. Whitman, Supervisor

Walter L. Fast

David W. Hoffman

Sean M. Kerwin

Yan (Jessie) Zhang

Structure-Based Mechanism of Hydratase-Aldolases in the Type I Aldolase Superfamily and Structures of Tautomerase Superfamily Members

by

Jake Allen LeVieux, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2017

Dedication

To my family in San Diego. To my Mom and Dad, who have been faithful in every phase of my life. You have been a model of integrity, respect, and authenticity, and for that I am forever thankful. Everything I do, including the work in this dissertation, is possible because of your care. To my brothers, Trevor and Lucas. You have changed me more than I think you know. I can see myself in both of you and this gives me a sense of self-compassion and confidence because it reminds me I am not ever alone. To my nephew, Jude, who has helped me understand human instinct at seven years old. To my grandmother, Baba, who has been so present throughout my life, and is a true example of strength and poise. To my aunt Cheri and my cousins Megan, Michelle, and Emily. To the families of John and Lyn Allen, and Jim and Peggy Allen. To my grandfather Rex Allen, who lives in us all.

To my local family in Austin — my grandparents Larry and Nana, my uncle Ted, my aunt Amanda, and cousins Kayla, Niki, and Hannah. Your support has been constant throughout my highs and lows during graduate school. The smiles we share make me feel at home. The weekly family dinners (or brunches) are precious to me, and I am a better person because of them. To Elizabeth and Larry Wilson and their daughters Helen (bff) and Margaret. To the family of James and Ryann LeVieux, who make the best of every moment. Plus, their iPad was used to remotely collect crystal diffraction data on a manly beach camping trip. To the aunts, uncles, and cousins that I have connected with at cousin's camp and family reunions.

Acknowledgements

The work in this dissertation is supported by the National Institutes of Health Grant GM-41239 (CPW) and the Robert A. Welch Foundation. I thank my advisor, Dr. Christian Whitman, for his guidance during my time as a graduate student. Because of you, I have enjoyed such freedom in my pursuit of discovery in science that has been exhilarating. Simultaneously, you have also kept me (as best you can) out of the weeds of interesting, but irrelevant science. I thank Dr. Tyler Stack for introducing me to Dr. Whitman, his dedication in managing our lab, and for the intellectual and personal conversations we have had. It was nice to have someone that would go to seminar with me. I think Dr. William Johnson Jr. for his jokes, good spirit, and wizardry in chemistry. I thank Kaci Erwin, the vicious athlete, for letting me borrow (take) enzymes and for playing Christmas music in lab. I thank Dr. Bert-Jan Baas for his mentorship over the past few years. I felt so alive discussing Fused 4-OT and its bizarre asymmetry. We stumbled upon a configuration I wondered might be possible when browsing the symmetry of protein homomultimers in a Biochemistry textbook. I thank Emily Lancaster for her clear communication and unifying spirit in our lab. I thank Dr. Jamison Huddleston for sharing his expertise and for giving me an additional opportunity to determine protein structures. I thank Samuel Guimarães for our conversations and for giving me a plasmid encoding an enzyme studied in his lab — that is trust. I thank Ingrid Johnson for her help with purification of PhdJ mutants.

I thank Dr. Jessie Zhang and her lab members for their perspective in the world of crystallography. The weekly lab meetings were a pleasure. I thank Dr. Wenzong Li, for showing me the ropes in crystallography from crystal screening to MAD phasing and refinement. I thank Dr. Wupeng Yan for his technical advice when I struggle with structure determination. I thank Brenda Medellin for our discussions and for her positivity and courage. I thank Wendy Matthews for being exceptionally authentic and making sure I have all I need in lab. I thank Dr. Art Monzingo for his earnest support and advice in all areas related to crystallography.

I thank Dr. Ben Liu and the members of his lab for their positive influence in my education. I thank Dr. Mark Ruszczycky for his wit and honest feedback. I thank Dr. Hak Jong Kim his efforts in the study of AprD4 — a remarkable enzyme. I thank Dr. Reid McCarty (deceased) for initiating the AprD4 project with me. I thank Dr. Sean Kerwin and Dr. Kenneth Johnson for the worthwhile experiences in their laboratories during my rotations. I thank Dr. Walter Fast for his wisdom in my qualifying exam, my committee meetings, and in this period as I prepare for my defense.

I thank all of the professors that I have assisted in teaching including Dr. David Hoffman, Dr. Jeffery Barrick, Dr. Gail Grabner, Dr. Jon Robertus, and Dr. Adrian Keatinge-Clay. I believe I am a more balanced scientist because of the diverse material I was exposed to in their courses. I might have learned a thing or two about teaching, as well.

Structure-Based Mechanism of Hydratase-Aldolases in the Type I Aldolase Superfamily and Structures of Tautomerase Superfamily Members

Jake Allen LeVieux, Ph.D.

The University of Texas at Austin, 2017

Supervisor: Christian P. Whitman

Polycyclic Aromatic Hydrocarbons (PAHs) are composed of multiple benzene-like rings and their bacterial catabolism has potential utility for bioremediation. In the breakdown of each ring, oxygenation is followed by ring cleavage and side chain removal, the latter of which is catalyzed by hydratase-aldolases. These enzymes belong to the *N*-acetylneuraminate lyase (NAL) subgroup of the Type I aldolase superfamily. NahE and PhdJ are hydratase-aldolases that process benzylidenepyruvate compounds in PAH degradative pathways from *Pseudomonas putida* G7 and *Mycobacterium vanbaalenii* PYR-1, respectively. Crystal structures of these enzymes in liganded and unliganded states are reported here and reveal new details about their catalytic mechanisms. Stabilization of the developing hydroxide anion during hydration could involve the amide nitrogen of Asn-157 (NahE numbering), conserved in hydratase-aldolases, but not other NAL subgroup members. For NahE, Tyr-155 might act as the general base for addition,

but the identity of the base is less certain for PhdJ. The crystal structure of PhdG reveals a carbinolamine intermediate hydrogen bonding to Tyr-152, consistent with the participation of the residue in Schiff base formation.

In addition, this dissertation describes structures within the tautomerase superfamily (TSF). The enzymes reported are from 4-oxalocrotonate tautomerase (4-OT), *cis*-*3*-chloroacrylic acid dehalogenase (*cis*-CaaD), and malonate semialdehyde decarboxylase (MSAD) subgroups. 4-OT catalyzes the ketonization of 2-hydroxymuconate (2-HM) in the *meta*-fission pathway. *cis*-CaaD and MSAD respectively catalyze dehalogenation and decarboxylation reactions in the catabolic 1,3-dichloropropene pathway, respectively. TSF members have a catalytic Pro-1 and share a common β-α-β fold. TSF members designated Ps01740, Pt0534, and Fused 4-OT have sequence similarity with the *cis*-CaaD and 4-OT subgroups and may provide insight into gene duplication and fusion events in the TSF. These enzymes have tautomerase activity and low or absent *cis*-CaaD activity. Liganded and unliganded structures suggest a structural basis for their activation. The trimeric structure of Fused 4-OT is unusual in that one monomer is inverted relative to the others. Mechanisms for 2-HM tautomerization are proposed based on a 2-HM-bound structure and kinetic parameters of mutants. Crystal structures for the MSAD homologues BP4401 and YusQ were obtained, and provide clues about their activities.

Table of Contents

List of Tables xiii

List of Figures xv

List of Schemes xx

Chapter 1: The Roles of the Hydratase-Aldolases in the Degradation of Polycyclic Aromatic Hydrocarbons 1

1.1 Bioremediation of Polycyclic Aromatic Hydrocarbons 1

1.1.1 Properties, Sources, and Toxicity of Polycyclic Aromatic Hydrocarbons 1

1.1.2 Bioremediation of PAHs 3

1.2 NahE, A Hydratase-Aldolase in the Upper Pathway of Naphthalene Degradation in *Pseudomonas putida* G7 4

1.2.1 *Pseudomonas putida* G7 and the NAH7 Plasmid 4

1.2.2 The Upper Pathway for Naphthalene Degradation in *P. putida* G7 6

1.2.3 The Hydratase-Aldolase (NahE) 10

1.3 PAH Metabolic Network in *Mycobacterium vanbaalenii* PYR-1 16

1.3.1 PAHs Degraded and Gene Clustering 17

1.3.2 Enzymes of the PAH Metabolic Network 19

1.3.3 Epistasis and Pleiotropy 22

1.3.4 Phenanthrene Catabolism in *M. vanbaalenii* PYR-1 24

1.3.5 The Hydratase-Aldolases PhdJ and PhdG 26

1.4 The *N*-Acetylneuraminate Lyase Sub-Family of Type I Aldolases 28

1.4.1 Type I Aldolases 28

1.4.2 The *N*-Acetylneuraminate Lyase (NAL) Subgroup 28

1.4.3 *N*-Acetylneuraminate Lyase 31

1.4.4 Dihydrodipicolinate Synthase 34

1.4.5 D-2-Keto-3-deoxygluconate Aldolase 37

1.4.6 Keto-deoxy-D-galactarate Dehydratase 38

Chapter 2: The Hydratase-Adolase Homologues NahE, PhdJ, and PhdG.....42
2.1 Introduction.....42
2.2 Experimental Procedures.....42
2.2.1 Preparation of NahE, PhdJ, and PhdG.....44
2.2.2 Synthesis.....52
2.2.3 Steady-State Kinetics.....55
2.2.4 ^{1}H NMR Spectrometric Time Course for Reactions Catalyzed by NahE, NahD, PhdJ, and PhdG.....61
2.2.5 UV/Vis Absorbance and Mass Spectrometry of Modified Hydratase-Aldolases.....63
2.2.6 Dynamic Light Scattering Analysis of NahE, PhdJ, and PhdG..65
2.2.7 Crystallization, Soaking, and Structure Determination of NahE 65
2.2.8 Crystallization, Soaking, and Structure Determination of PhdJ .67
2.2.9 Crystallization, Soaking, and Structure Determination of PhdG 68
2.3 Results.....69
2.3.1 Purification and Enzyme Activity of NahE, PhdJ, PhdG.....69
2.3.2 Steady-State Kinetic Parameters for NahE, PhdJ, PhdJ Mutants, and PhdG.....77
2.3.3 Steady-State Kinetic Parameters of NahE and PhdG Using Pyruvate and an Aldehyde.....81
2.3.4 ^{1}H NMR Spectrometric Analysis of Reactions Catalyzed by NahE, NahD, PhdJ, and PhdG.....83
2.3.5 Modification of NahE, PhdJ, and PhdG by Various Aldehydes and α-Keto Acids.....86
2.3.6 Dynamic Light Scattering.....94
2.3.7 Crystal Structures of NahE.....95
2.3.8 Crystal Structures of PhdJ.....105
2.3.9 Crystal Structures of PhdG.....114
2.4 Discussion.....119
2.4.1 Hydratase-Aldolases in the NAL Subgroup.....119
2.4.2 The Mechanism of NahE Using HBP.....124
2.4.3 The Mechanism of PhdJ Using CBP.....129

2.4.4 The S278N and D282E PhdJ Mutants 137
2.4.5 Structure and Function of PhdG 138
2.4.6 Biological Roles for NahE, PhdJ, and PhdG 142

Chapter 3: The 4-Oxalocrotonate Tautomerase, *cis*-3-Chloroacrylic Acid Dehalogenase, and Malonate Semialdehyde Decarboxylase Subgroups of the Tautomerase Superfamily 144
3.1 4-Oxalocrotonate Tautomerase 145
3.2 1,3-Dichloropropene Catabolic Pathway 149
3.3 *trans*-3-Chloroacrylic Acid Dehalogenase 150
3.4 *cis*-3-Chloroacrylic Acid Dehalogenase 152
3.5 Malonate Semialdehyde Decarboxylase 157

Chapter 4: Enzymes Linking 4-Oxalocrotonate Tautomerase and *cis*-3-Chloroacrylic Acid Dehalogenase Subgroups 161
4.1 Introduction 161
4.2 Experimental Procedures 163
4.3 Results 165
4.3.1 Crystal Structure of Ps01740 and Ps01740-2OP 165
4.3.2 Crystal Structure of Pt0534-2OP 170
4.3.3 Crystal Structure of Fused 4-OT 173
4.3.4 Crystal Structures of F4OT-2HM, F4OT-2OP, and F4OT-PP .181
4.4 Discussion 190
4.4.1 Ps01740 Structure and Function 191
4.4.2 Pt0534 Structure and Function 196
4.4.3 Fused 4-OT Structure and Function 200
4.4.4 2-HM Tautomerization in Interface A and B of Fused 4-OT ...204
4.4.5 Structure and Function Relationships of Linkers 209

Chapter 5: The Malonate Semialdehyde Decarboxylase Homologues BP4401 and YusQ 211
5.1 Introduction 211
5.2 Experimental Procedures 212
5.3 Results 213

5.4 Discussion 221

References 225

List of Tables

Table 2-1 Steady-state parameters of NahE and PhdJ with HBP, CBP and BP. ..79

Table 2-2 Kinetic parameters of the S278N and D282E PhdJ mutant using HBP, CBP, and BP. 80

Table 2-3 Steady-state parameters of PhdG using BP. 81

Table 2-4 NahE and PhdG kinetic constants with aldehyde (ald, specifically salicylaldehyde and cinnnamaldehyde, respectively) and pyruvate. 83

Table 2-5 ESI-MS signal peak values with modification of NahE using NaCNBH3. 90

Table 2-6 ESI-MS signal peak values with modification of K183A NahE using NaCNBH3. 91

Table 2-7 ESI-MS signal peak values with modification of PhdJ using CBP, CBA, and pyruvate in the presence of $NaCNBH_3$. 93

Table 2-8 ESI-MS signal peak values with modification of PhdG using various compounds in the presence of $NaCNBH_3$. 94

Table 2-9 Crystallographic data and refinement statistics for NahE-SeM, NahE-HN, and NahE-SP. 101

Table 2-10 Crystallographic data and refinement statistics for NahE-Sal, NahE-BP, and NahE-BPy. 104

Table 2-11 Crystallographic data and refinement statistics for PhdJ, PhdJ-CBP, PhdJ-CN, and PhdJ-CBA. 113

Table 2-12 Crystallographic data and refinement statistics for PhdG and PhdG-Pyr. 118

Table 4-1 Sequence identity of 4-OT and *cis*-CaaD with Ps01740, Pt0534, and Fused 4-OT. 162

Table 4-2 Crystallographic data and refinement statistics for Ps01740 and Ps01740-2OP. 169

Table 4-3 Crystallographic data and refinement statistics for Pt0534-2OP. 172

Table 4-4 Crystallographic data and refinement statistics for F4OT. 180

Table 4-5 Crystallographic data and refinement statistics for F4OT-2HM, F4OT-2OP, and F4OT-PP. 189

Table 4-6 Steady-state parameters for Ps01740, Pt0534, Fused 4-OT (Baas BJ, *In Preparation*), *cis*-CaaD, and 4-OT. 203

Table 4-7 Steady-state Kinetic Parameters of Fused 4-OT Mutants (Baas BJ, *In Preparation*). 207

Table 5-1 Crystallographic Data and Refinement statistics for BP4401, BP4401-2OP, BP4401-BrP, YusQ, and YusQ-BrP. 220

List of Figures

Figure 1-1 Representative PAHs.....1

Figure 1-2 Modification of DNA by dihydrodiol epoxide benzo[*a*]pyrene.....2

Figure 1-3 Operon for the upper naphthalene degradation pathway on the NAH7 plasmid.....5

Figure 1-4 Structures of **A)** HBP: R=OH, BP: R=H, CBP: R=COO^- **B)** *m*-HBP **C)** 2,6-dihydroxybenzylidenepyruvate **D)** 2,4-dihydroxybenzylidenepyruvate.....12

Figure 1-5 Structures of **A)** salicyaldehyde: R=OH, benzaldehyde: R=H **B)** 3-phenylpropionaldehyde **C)** cinnamaldehyde. 14

Figure 1-6 PAHs mineralized by *Mycobacterium vanbaalenii* PYR-1. 18

Figure 1-7 Bay and K regions of several PAHs.....21

Figure 1-8 Genomic context of PhdJ and PhdG. 27

Figure 1-9 The crystal structure of *E. coli* NAL (2.2 Å resolution, PDB entry 1NAL). 30

Figure 1-10 The crystal structure of *Haemophilus influenza* NAL (1.6 Å resolution, PDB entry 1F7B) liganded with the C-4 *gem*-diol of substrate analogue *N*-acetyl-4-oxoneuraminate.....33

Figure 1-11 The crystal structure of the *E. coli* dihydrodipicolinate synthase covalently bound with condensation product of pyruvate and succinate (2.3 Å resolution, PDB entry 4EOU).....36

Figure 1-12 Active site of the KDGA showing D-2-keto-3-deoxygluconate covalently bound at Lys-155 (2.1 Å resolution, PDB entry 1W3N). 38

Figure 1-13 Active site of the KDGDH covalentely bound with 2-oxoadipic acid (1.9 Å resolution, PDB entry 1EC8). 39

Figure 2-1 Purified NahE and absorbance spectra of HBP in the presence of NahE, PhdJ, and PhdG. 70

Figure 2-2 Isomerization of HBP over time and at basic pH. 71

Figure 2-3 A) Initial rates of NahE using HBP (98 μM) with salicyaldehyde in 50 mM sodium phosphate buffer (pH 7.0). **B)** Rates obtained with variable HBP and salicylaldehyde (procedure is in **Section 2.2.3**). 72

Figure 2-4 Spectra showing the reaction of CBP in the presence of NahE, PhdJ, and PhdG. 74

Figure 2-5 BP production/consumption by NahE, PhdJ, and PhdG 75

Figure 2-6 Absorbance spectra during conversion of *trans*-cinnamaldehyde and pyruvate to 2-oxo-6-phenyl-3,5-hexadienoate by NahE, PhdJ, and PhdG. 76

Figure 2-7 Initial rates for NahE with HBP reveal the enzyme is inhibited by substrate and possesses a low K_M value (20-fold less than K_I). 78

Figure 2-8 3-D plot of k_{obs} vs. conc. and fits to bisubstrate reactions. 82

Figure 2-9 ^{1}H NMR spectra of NahE and NahD using HBP and HCCA. 85

Figure 2-10 UV/Vis spectra of NahE and aldehyde substrates/products. 87

Figure 2-11 UV/Vis spectra of PhdJ and aldehyde substrates/products. 89

Figure 2-12 ESI-MS/MS spectrum of fragments from modified NahE incubated with pyruvate and $NaCNBH_3$ 92

Figure 2-13 The 1.9 Å resolution crystal structure of NahE-SeM. 96

Figure 2-14 Active site of the 2.1 Å resolution NahE-HN crystal structure. 99

Figure 2-15 Active site of the 2.3 Å resolution crystal structure of NahE-SP. 100

Figure 2-16 Active site of the 2.9 Å resolution crystal structure of NahE-Sal. ..102

Figure 2-17 The 1.9 Å resolution crystal structures of NahE-BP and NahE-BPy103

Figure 2-18 The 2.1 Å resolution crystal structure of PhdJ. 106

Figure 2-19 Active site of the 2.0 Å resolution PhdJ-CBP crystal structure**.** 109

Figure 2-20 Active site of the 2.0 Å resolution PhdJ-CN crystal structure. 110

Figure 2-21 Active site of PhdJ-CBP with pyruvate and CBA.....111

Figure 2-22 Active site of the 2.2 Å resolution PhdJ-CBA crystal structure with $2F_o$-F_c electron density map (0.5 σ) of surrounding CBA and the Schiff base linkage to Lys-180. 112

Figure 2-23 The 2.7 Å resolution crystal structure of PhdG.....115

Figure 2-24 Active site of the 1.9 Å resolution PhdG-Pyr crystal structure. 117

Figure 2-25 Sequence alignment of representative NAL subgroup members. ...121

Figure 2-26 The helix α10 unique to hydratase-aldolases in the NAL subgroup.123

Figure 2-27 Superposition of NahE crystal structures. 125

Figure 2-28 Superposition of PhdJ crystal stuctures.....131

Figure 2-29 Candidates for water activation and proton donation during hydration of CBP by PhdJ. 134

Figure 2-30 Superposition of NahE-HN (green) and PhdJ-CN (slate) active sites.138

Figure 2-31 Superposition of PhdG and PhdG-Pyr. 140

Figure 2-32 Superposition of NahE (forest), PhdJ (slate), and PhdG (light teal).141

Figure 3-1 Assemblies of long and short monomers in the TSF. 145

Figure 3-2 Crystal structure of 4-OT (1.9 Å resolution, PDB entry 1OTF). 147

Figure 3-3 Crystal structure of Native *cis*-CaaD (2.0 Å resolution, PDB entry 3MF8) 153

Figure 3-4 Crystal structure of MSAD (1.9 Å resolution, PDB entry 2AAG)**.** ..158

Figure 4-1 The 2.6 Å resolution crystal structure of Ps01740. 166

Figure 4-2 Active sites of Ps01740 (2.6 Å resolution) and Ps01740-2OP (2.5 Å resolution). 168

Figure 4-3 The 2.0 Å resolution crystal structure of Pt0534-2OP. 171

Figure 4-4 Monomer of the 1.5 Å resolution Fused 4-OT (F4OT) crystal structure. 173

Figure 4-5 The 1.5 Å resolution F4OT crystal structure.....175

Figure 4-6 Comparison of F4OT monomers.....177

Figure 4-7 Comparison of active sites in F4OT.....179

Figure 4-8 The 1.9 Å resolution structure of F4OT-2HM shown with $2F_0$-F_c map contoured at 0.5 σ. 182

Figure 4-9 The 2.3 Å resolution F4OT-2OP crystal structure. 184

Figure 4-10 Superposition of **A)** F4OT-2HM and **B)** F4OT-2OP active sites. ..186

Figure 4-11 Active site of the 2.7 Å resolution F4OT-PP crystal structure.....188

Figure 4-12 Sequence alignment of enzymes of the 4-OT and *cis*-CaaD subgroups. 191

Figure 4-13 Superposition of Ps01740, *cis*-CaaD, and 4-OT. 193

Figure 4-14 Superposition of Ps01740-2OP and inactivated homologues. 196

Figure 4-15 Superposition of Pt0534-2OP with 4-OT and *cis*-CaaD. 198

Figure 4-16 Alignment of Pt0534 (deep teal and violet purple) and 4-OT labeled with 2-oxo-3-penteonate adduct (orange, PDB entry 1BJP).....200

Figure 4-17 Superposition of F4OT with 4-OT and *cis*-CaaD. 201

Figure 4-18 Comparison of phenylpyruvate bound to **A)** Fused 4-OT (F4OT-PP) and **B)** heterohexamer 4-OT (Burks *et al.*, 2010). 204

Figure 4-19 Active site superposition of F4OT-2HM (orange and deep teal) and 4-OT (limon, PDB entry 1OTF).....205

Figure 5-1 The 2.2 Å resolution BP4401 crystal structure and the 2.1 Å resolution YusQ crystal structure.....214

Figure 5-2 Active sites of the 1.9 Å resolution structure of BP4401 inactivated by 2-OP. 216

Figure 5-3 Active sites of the 1.8 Å resolution BP4401-BrP structure and the 2.8 Å YusQ-BrP structure.....219

Figure 5-4 Superposition of MSAD (deep salmon, PDB entry 2AAG), BP4401 (deep teal), and YusQ (forest) monomers.....221

Figure 5-5 Superposition of MSAD with 3-oxopropanoate adduct (deep salmon, PDB entry 2AAL) and YusQ-BrP with 2-oxopropanoate adduct (lime green). 223

Figure 5-6 Sequence alignment of MSAD, FG41, BP4401, and YusQ.....224

List of Schemes

Scheme 1-1 NahE conversion of HBP to salicyaldehyde and pyruvate.....4

Scheme 1-2 Degradation of naphthalene by *Pseudomonas putida* G7 NAH7 plasmid. 8

Scheme 1-3 NahD-catalyzed isomerization of HCCA to HBP. 9

Scheme 1-4 Proposed catalytic cycle for NahE using HBP to produce salicylaldehyde (sal) and pyruvate.....11

Scheme 1-5 A) Hydration of HBP via a *ortho*-quinone methide intermediate. **B)** Decarboxylation of 4-hydroxycinnamate via a *para*-quinone methide.....13

Scheme 1-6 Evidence for stability of 4-hydroxy intermediates and NahE-catalyzed dehydration in condensation reactions.....16

Scheme 1-7 General cycle for the degradation of a single ring of a PAH. 20

Scheme 1-8 Phenanthrene catabolic pathway via 3,4-dihydroxyphenanthrene. ...25

Scheme 1-9 Aldol mechanism for Class I Aldolases. 28

Scheme 1-10 Mechanism for NAL-catalyzed conversion of N-acetyneuraminate (sialic acid) to acetyl-D-mannosamine and pyruvate. 32

Scheme 1-11 DHDPS-catalyzed condensation of pyruvate and (*S*)-aspartate semialdehyde hydrate.....35

Scheme 1-12 The KDGA-catalyzed conversion of D-2-keto-3-deoxygluconate or D-2-keto-3-deoxygalactonate to D-glyceraldehyde and pyruvate. 37

Scheme 1-13 Proposed mechanism for KDGDH. 41

Scheme 2-1 Synthesis of HBP (R=OH) and CBP ($R=CO_2^-$), starting with salicylaldehyde (R=OH) or CBA ($R=CO_2^-$). 53

Scheme 2-2 Proposed mechanism for NahE using HBP. 129

Scheme 2-3 Proposed mechanism of PhdJ using CBP. 136

Scheme 3-1 Role of 4-OT in the *meta*-fission pathway. 146

Scheme 3-2 Mechanism of 2-HM tautomerization by 4-OT. 148

Scheme 3-3 Inactivation of 4-OT by 2-OP. 149

Scheme 3-4 Degradation of racemic 1,3-dichloropropene. 150

Scheme 3-5 Mechanism for CaaD using *trans*-3-chloroacrylate. 151

Scheme 3-6 Mechanism for *cis*-CaaD using *cis*-3-chloroacrylate. 154

Scheme 3-7 Hydration mechanism for 2-OP. 155

Scheme 3-8 Ketonization of phenylenolpyruvate. 156

Scheme 3-9 Covalent labeling of Pro-1 (prolyl nitrogen) after incubation with 3-bromopropiolate. 157

Scheme 3-10 Mechanism for decarboxylation of malonate semialdehyde by MSAD. 159

Scheme 4-1 A) Direct and **B)** covalent hydration mechanism of Ps01740 using 2-OP. 195

Scheme 4-2 Proposed mechanisms for **A)** 4-OT **B)** Fused 4-OT (Interface A) **C)** Fused 4-OT (Interface B). 208

Scheme 5-1 Possible BP4401 and YusQ modifications at Pro-1 (prolyl nitrogen) using 3-bromopropiolate. 217

Chapter 1: The Roles of the Hydratase-Aldolases in the Degradation of Polycyclic Aromatic Hydrocarbons

1.1 Bioremediation of Polycyclic Aromatic Hydrocarbons

1.1.1 Properties, Sources, and Toxicity of Polycyclic Aromatic Hydrocarbons

Polycyclic aromatic hydrocarbons (PAHs) are multi-ring structures with conjugated π systems. Low molecular weight (LMW) PAHs, such as naphthalene or phenanthrene, are made of three or fewer rings, but high molecular weight (HMW) PAHs contain four or more rings making them less soluble (Wild & Jones, 1995) and resistant to degradation (Kuśmierz *et al.*, 2016).

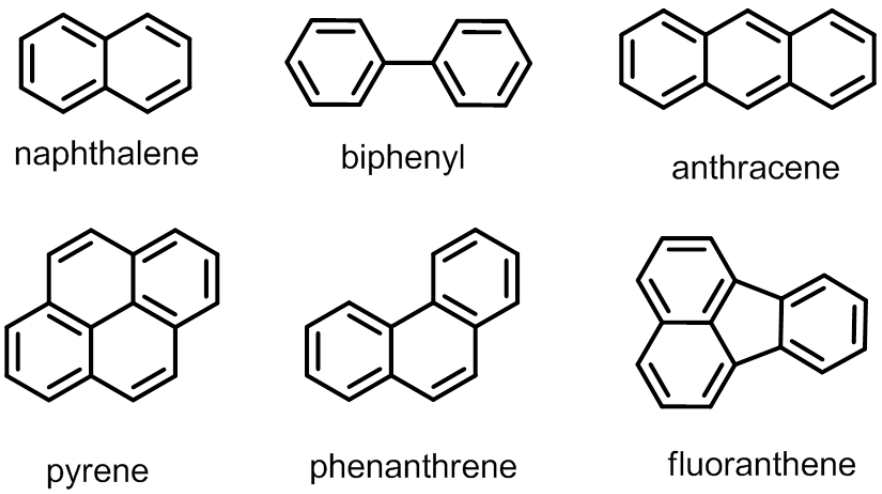


Figure 1-1 Representative PAHs.

PAHs are generated naturally and anthropologically in a process often involving heat, high pressure, and the lack of oxygen (Douben, 2003). Anthropogenic production of PAHs occurs through combustion of fossil fuels or wood, pyrolysis (coal distillation), and oil leakage. Natural sources include forest fires, volcanos, and petroleum seeps (European Commission, 2002). LMW PAHs can also be produced biologically — naphthalene has been shown to be produced in *Magnolia* flowers (Azuma *et al.*, 1996), and in an endophytic fungus (Daisy *et al.*, 2002).

Sixteen PAHs have been targeted by the EPA as priority pollutants and are the focus of environmental research (Ranc *et al.*, 2016). Dihydrodiol epoxides such as (+)-benzo[*a*]pyrene-7,8-dihydrodiol-9,10-epoxide (**Figure 1-2A**) are vulnerable to attack by exocyclic amines of purine bases (**Figure 1-2B),** and are associated with inactivation of tumor suppressor genes and activation of proto-oncongenes (Yeh *et al.*, 1995, Hecht, 1999).

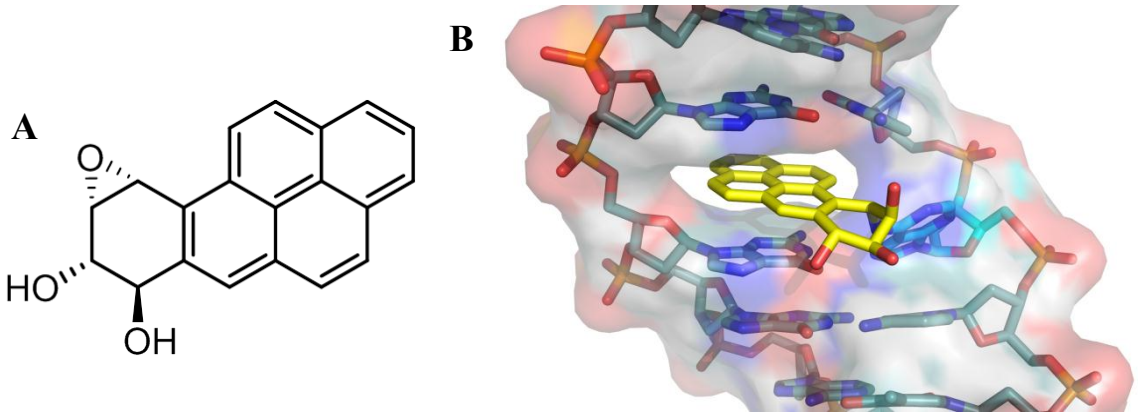


Figure 1-2 Modification of DNA by dihydrodiol epoxide benzo[*a*]pyrene.

A) (+)-benzo[*a*]pyrene-7,8-dihydrodiol-9,10-epoxide. **B)** NMR solution structure of N^6-amino modification of deoxyadenosine by the structure shown in **A** (PDB entry 1DXA).

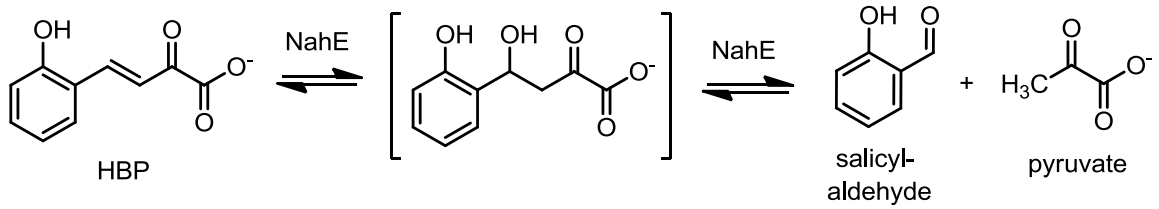
1.1.2 Bioremediation of PAHs

For almost a century, bacteria have been known to degrade PAHs, either cometabolically or as a sole carbon source (Peng *et al.*, 2008). Tausson and coworkers isolated three bacterial strains that degraded phenanthrene, anthracene, and other PAHs (Tausson, 1928). It was later discovered that bacteria responsible for the degradation of LMW PAHs are ubiquitous (Sisler & Zobell, 1947, Davies & Evans, 1964). The report of a bacterium that could degrade fluoranthene and pyrene (Heitkamp & Cerniglia, 1988) was the first example of microbial degradation of HMW PAHs (four or more rings, **Section 1.3**).

Bacterial bioremediation is used for removing PAHs and has aided in the clean-up of large oil spills (Atlas & Hazen, 2011). Relatively few organisms are capable of degrading HMW PAHs, such as benzo[*a*]pyrene in comparison to the number of organisms able to degrade LMW PAHs. The large number of organisms able to degrade LMW PAHs sharply contrast with the relatively small number able to degrade HMW PAHs such as benzo[*a*]pyrene. This discrepancy in the ability of microbes to break down different size PAHs, in conjunction with the low water solubility of HMW PAHs, causes these compounds to persis in the environment at a higher rate than LMW PAHs (Boldrin *et al.*, 1993, Tiehm, 1994, Juhasz & Naidu, 2000). An understanding of degradation pathways could help engineer microbes with improved catabolic abilities.

1.2 NahE, A Hydratase-Aldolase in the Upper Pathway of Naphthalene Degradation in *Pseudomonas putida* G7

Napthalene degradation in *P. putida* G7 is the most extensively studied PAH catabolic pathway and involves operons for the upper and lower pathways on the NAH7 plasmid. In the upper pathway, *trans*-*o*-hydroxybenzylidenepyruvate (**HBP**) hydratase-aldolase (**NahE**) catalyzes a hydration reaction followed by a retro-aldol reaction. Previous studies of NahE suggest that both steps of the reaction are enzyme-catalyzed, and that the *o*-hydroxyl of the aromatic substrate is important for binding and HBP turnover.



Scheme 1-1 NahE conversion of HBP to salicyaldehyde and pyruvate.

1.2.1 *Pseudomonas putida* G7 and the NAH7 Plasmid

Study of naphthalene degradation has a long history, and has recently focused on a salicylate-induced pathway encoded in a transmissible NAH7 plasmid from

Pseudomonas putida G7. Discoveries in this pathway began before DNA sequencing was developed. Bacterial growth on naphthalene-mineral salts was described in 1928, and intermediates including 1,2-dihydrodihydroxynaphthalene and salicylic acid were isolated in the 1940s and 1950s (Davies & Evans, 1964). The NAH7 plasmid contains operons for the upper pathway (**Figure 1-3**, conversion of naphthalene to salicylate) and lower pathway (conversion of salicylate to pyruvate and acetaldehyde), which were likely obtained through horizontal gene transfer (Sota *et al.*, 2006). Expression of both lower and upper pathway enzymes are induced in the presence of salicylate by NahR, a positive regulator (Chauhan *et al.*, 2008).

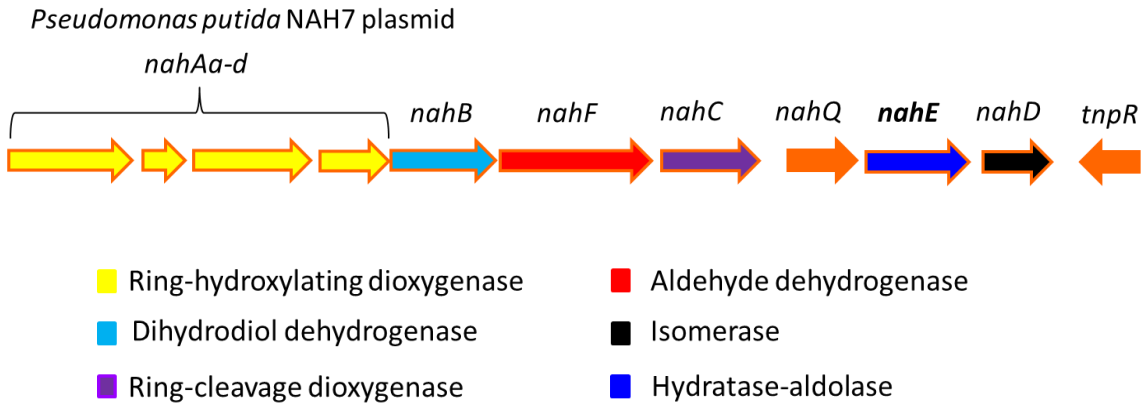


Figure 1-3 Operon for the upper naphthalene degradation pathway on the NAH7 plasmid.

Although the pathway for naphthalene degradation is established, NAH7 gene products may facilitate the degradation of large PAHs as well. Anthracene and phenanthrene degradation products such as naphthoic acid were detected from

Pseudomonas fluoresecens 5R, which contained a plasmid similar to NAH7 (Menn *et al.*, 1993). These intermediates suggest the known enzymes from the NAH7 plasmid tolerate larger substrates for degrading three-ring PAHs.

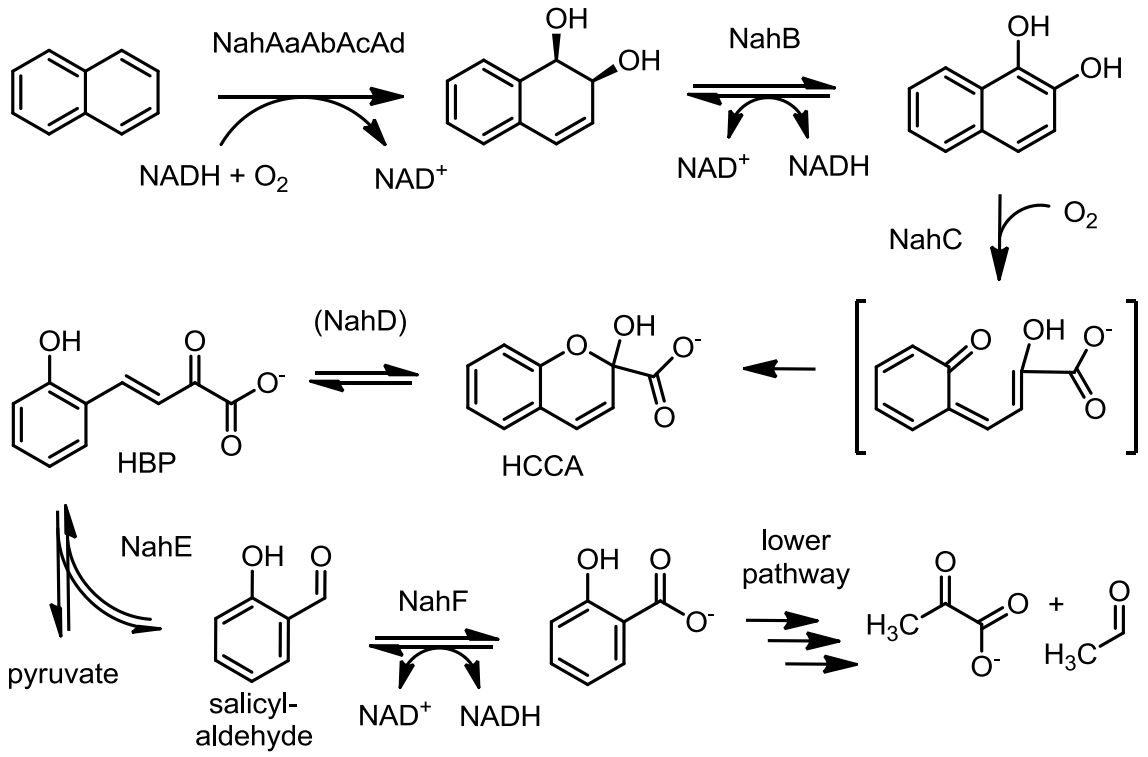
1.2.2 The Upper Pathway for Naphthalene Degradation in *P. putida* G7

Initial dihydroxylation of naphthalene (**Scheme 1-2**) is facilitated by a ring-hydroxylating oxygenase (RHO) system comprised of four gene products NahAaAbAcAd. Electrons from NADH are initially transferred to an iron-sulfur flavoprotein NahAa. This protein reduces NahAb, a Rieske ferredoxin, which finally transfers electrons to the Rieske center of naphthalene dioxygenase NahAcAd. The protein is of $\alpha_3\beta_3$ composition, where each α subunit contains a Rieske [2Fe-2S] center and a mononuclear iron (Karlsson *et al.*, 2003). Electron access to mononuclear iron is gated by a conserved aspartate to initiate side-on binding of O_2 to Fe^{2+} to form naphthalene *cis*-dihydrodiol.

Conversion of naphthalene 1,2-*cis*-dihydrodiol to 1,2-dihydroxynaphthalene is catalyzed by a NAD^+-dependent dehydrogenase NahB (Eaton & Chapman, 1992). This protein is likely nonspecific, because several homologues oxidize *cis*-dihydrodiols with one to three rings including napthhalene, anthracene, phenanthrene, and biphenyl (Patel & Gibson, 1976), though none were found to process *trans*-dihydrodiols.

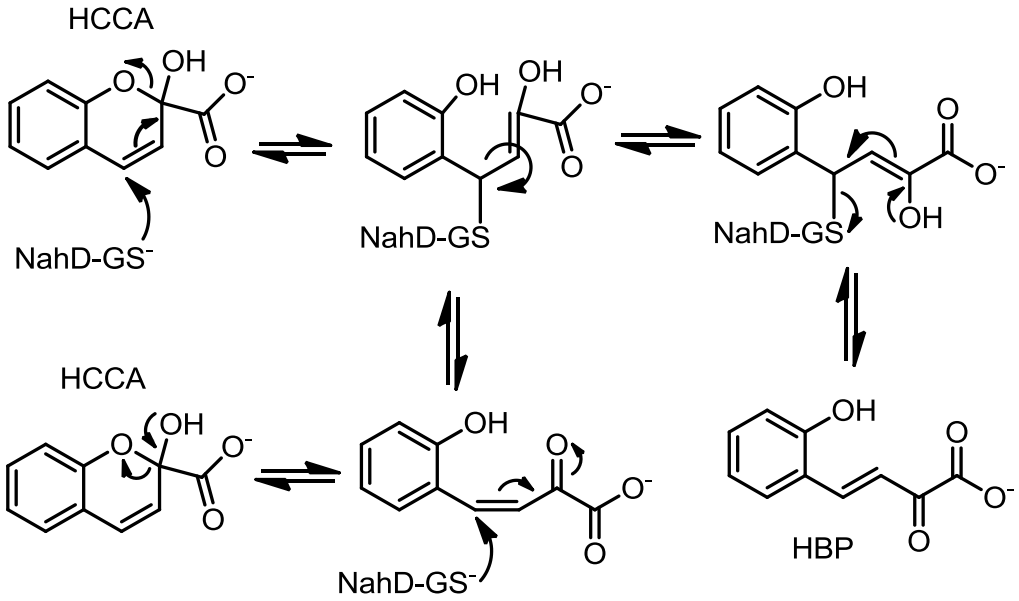
While *meta*-cleavage of 1,2-dihydroxylnaphthalene was apparent in 1965 (Evans *et al.*, 1965), identification and functional assignment of the enzymes involved in the

currently accepted pathway (**Scheme 1-2**) wasn't established until the enzymes were expressed heterologously in the 1990s (Eaton & Chapman, 1992). This report showed that 2-hydroxychromene-2-carboxylate (**HCCA**) was the initial product of ring cleavage by the dioxygenase NahC and that isomerization, assisted by NahD, occurred to form *trans*-*o*-hydroxybenzylidenepyruvate (**HBP**). Although HCCA is the observed product, a monocyclic species may be released from NahC prior to cyclization (Patel & Barnsley, 1980). Based on reactions catalyzed by extradiol dioxygenases, a quinone methide might be the first product by NahC prior to cyclization to HCCA (Vaillancourt *et al.*, 2006, Bugg & Ramaswamy, 2008). HCCA isomerizes to HBP, the substrate for HBP hydratase-aldolase (**NahE)**, non-enzymatically, or with the assistance of HCCA isomerase (**NahD**). NahE converts HBP to salicylaldehyde and pyruvate.



Scheme 1-2 Degradation of naphthalene by *Pseudomonas putida* G7 NAH7 plasmid.

The nonenzymatic isomerization of HCCA to HBP reaches an equilibrium mixture of 55% HCCA and 45% HBP in five hours (Eaton & Chapman, 1992). The interconversion is accelerated by NahD, a member of the glutathione transferase superfamily that uses tightly bound reduced glutathione (K_D of 5 nM) as a cofactor (Thompson *et al.*, 2007). Glutathione is proposed to attack the benzylic carbon for rotation around C-3 or C-4 of HBP or HCCA. During conversion of HCCA to HBP (forward direction in the pathway), attack could occur before or after ring opening of HCCA (**Scheme 1-3**).

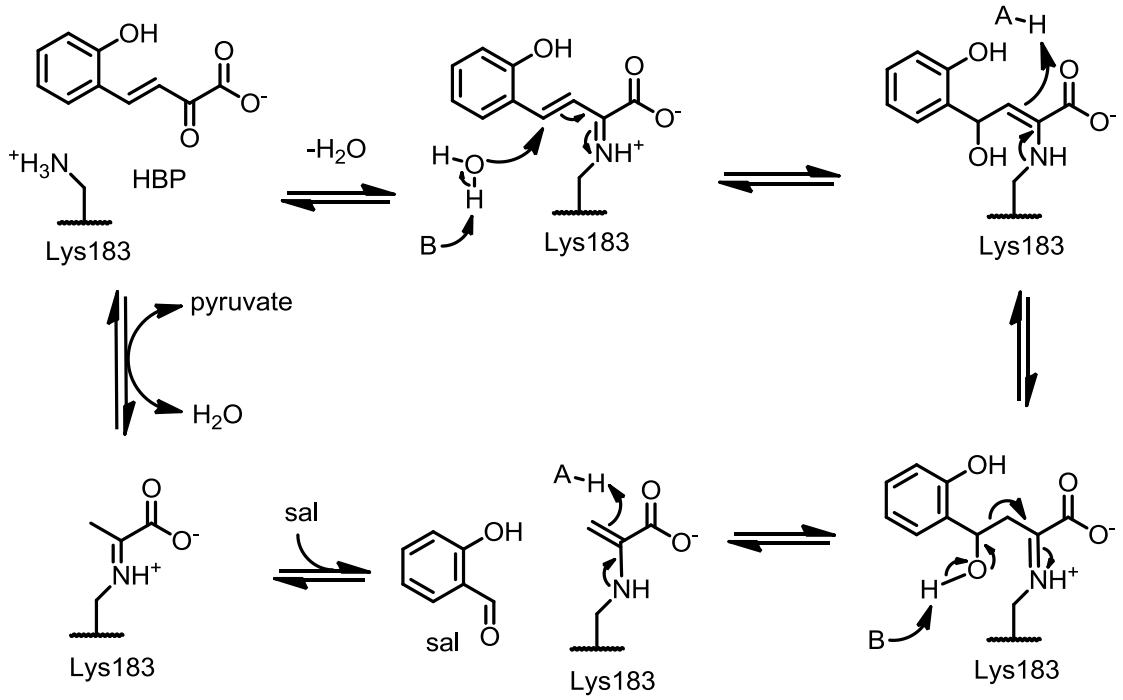


Scheme 1-3 Proposed mechanism for NahD-catalyzed isomerization of HCCA to HBP.

The final enzyme in the upper pathway is salicyaldehyde dehydrogenase (**NahF**) which oxidizes salicylaldehyde to salicylate using NAD^+ (Eaton & Chapman, 1992). Salicylic acid, the end product of the upper pathway, undergoes a decarboxylation reaction catalyzed by salicylate hydroxylase (NahG) encoded in the operon for the lower pathway. The remaining reactions of the lower pathway resemble meta fission pathways such as those for toluene and xylene, where enzymes are encoded on the TOL plasmid in *Pseudomonas putida* mt-2 (Yen & Gunsalus, 1982). Degradation of catechol to acetaldehyde and pyruvate in the NAH7 and TOL degradative pathways are identical, and the plasmids likely originated from a single ancestral sequence (Harayama *et al.*, 1987).

1.2.3 The Hydratase-Aldolase (NahE)

Hydratase-aldolases are understudied compared to other PAH degradative enzymes and other type I aldolases (**Section 1.4**). NahE is the most thoroughly characterized hydratase-aldolase involved in PAH degradation. For that reason it is the starting point for the work presented here. NahE forms a Schiff base with 4-phenyl-3,4-unsaturated α-keto acids, prefers *o*- and *m*- hydroxyl groups, and facilitates both hydration and retro-aldol cleavage without tautomerization using the *o*-hydroxyl group. A mechanism consistent with prior research of NahE is shown in **Scheme 1-4**. Lys-183 is conserved in the *N*-acetylneuraminate lyase subgroup (**Section 1.4**) and forms a Schiff base with HBP at C-2, consistent with the mechanisms for the enzymes which process α-keto acids within the subgroup. Water is then activated for addition at C-4 of HBP, with subsequent rearrangement to form the enamine. The enamine abstracts a proton from an active site group. Deprotonation of the C-4 hydroxyl allows for heterolytic cleavage of the intermediate and the formation of a pyruvoyl enamine, which tautomerizes to form the pyruvate Schiff base adduct. A conserved tyrosine (Tyr-155) in the *N*-acetyleneuroaminate Lyase subgroup (**Section 1.4**) is likely involved in Schiff Base formation and retro-aldol catalysis. Salicylaldehyde is also produced and the pyruvate imine is hydrolyzed to release pyruvate and the lysine.



Scheme 1-4 Proposed catalytic cycle for NahE using HBP to produce salicylaldehyde (sal) and pyruvate.

Further support for this mechanism comes from the work on a NahE homologue from *Pseudomonas vesicularis* DSM 6383. This homologue is inactivated by either HBP, salicylaldehyde, or pyruvate in the presence of $NaBH_4$ (Kuhm *et al.*, 1993), suggesting that it is a type 1 aldolase (**Section 1.4**) where a lysine residue forms a Schiff base with α-keto acid and aldehyde moieties. Reduction of the Schiff base produces an amino acid adduct, which cannot be hydrolyzed. The same homologue was shown to be inhibited by salicylaldehyde (product) with an IC_{50} of 15-20 μM without $NaBH_4$ (Kuhm *et al.*, 1993).

The *o*-hydroxyl group of HBP has also been shown to be important for NahE activity. With the exception of *trans*-*m*-hydroxybenzylidenepyruvate (***m*-HBP**, **Figure 1-**

4B), HBP analogues without *o*-hydroxyl substituents are processed at less than 2% of the rate compared to HBP (Eaton, 2000). The relative rates for *trans*-benzylidenepyruvate (**BP**) and *trans*-*o*-carboxybenzylidenepyruvate (**CBP**, **Figure 1-4A**) were 0.52% and 1.4% of that for HBP, respectively (Eaton, 2000). Consistent with this finding, conversion of BP by an NahE homologue from *Pseudomonas vesicularis* DSM 6383 showed a rate of 1% that of HBP (Kuhm *et al.*, 1993). However, *trans*-dihydroxybenzylidenepyruvate compounds with *o*-hydroxyl groups were processed efficiently. Crude extracts from *P. versivularis* BN6, *P. putida* KT2442 NAH7, and *Pseudomonas NCIB* 10535, all processed HBP, 2,4-dihydroxybenzylidenepyruvate, and 2,6-dihydroxybenzylidenepyruvate (**Figure 1-4C,D**) at similar rates. Based on these results, the conclusion is that hydroxyl groups at the *ortho* or *meta* positions of benzylidenepyruvate are important for binding and turnover.

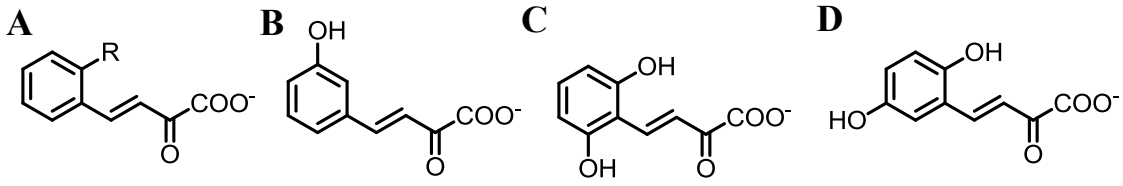
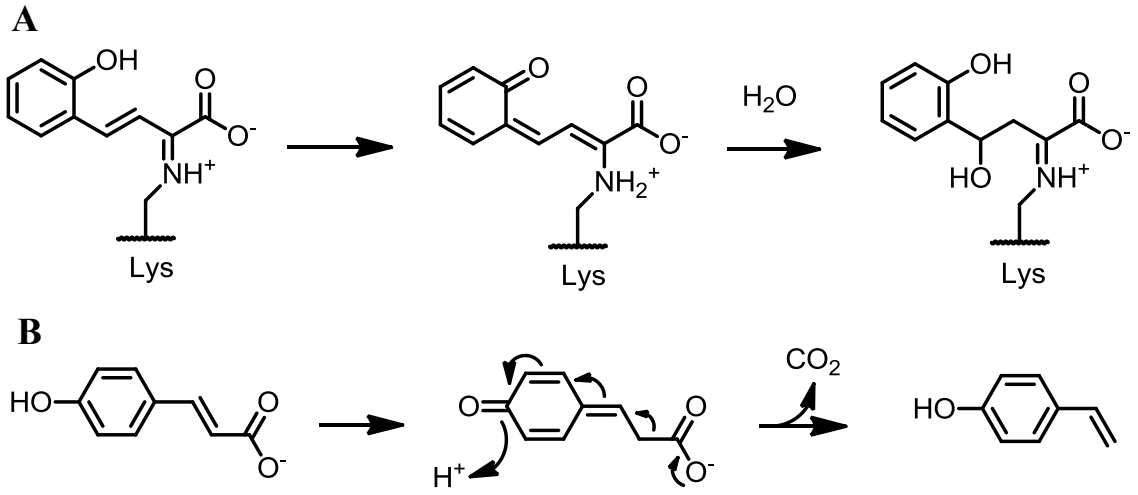


Figure 1-4 Structures of **A)** HBP: R=OH, BP: R=H, CBP: R=COO^- **B)** *m*-HBP **C)** 2,6-dihydroxybenzylidenepyruvate **D)** 2,4-dihydroxybenzylidenepyruvate.

The activation energy for hydration of a quinone methide intermediate could hypothetically be lower than for the HBP Schiff base (**Scheme 1-5A**). Precedence for a quinone methide intermediate comes from the 4-hydroxycinnamate decarboxylase-

catalyzed reaction (Ferrara *et al.*, 2011). In this reaction, the 4-hydroxycinnamate rearranges to form the quinone methide (**Scheme 1-5B**), allowing decarboxylation and the reformation of the 4-hydroxyl group (Hashidoko & Tahara, 1998). However, the NahE-catalyzed conversion of *trans-p*-hydroxybenzylidenepyruvate (***p*-HBP**) had 1% relative activity of wild-type, where *m*-HBP, an isomer that cannot form a quinone methide was processed at 75% of the wild-type rate (Eaton, 2000). These results suggest that a mechanism involving the aromatic ring is not operative.



Scheme 1-5 A) Hydration of HBP via a *ortho*-quinone methide intermediate. **B)** Decarboxylation of 4-hydroxycinnamate via a *para*-quinone methide.

In the reverse direction, NahE catalyzes aldol condensation reactions. In these reactions, pyruvate is the only donor known, but various aldehydes are accepted. The condensation reaction gives some insight about substrate preference in the forward direction. Over 30 aldehydes have been identified as substrates for NahE (Eaton, 2000).

The different rates for these substrates indicate that a nitro group at the *para* position accelerates the reaction in a *Pseudomonas fluorescens* N3 homologue (96% sequence identity). This suggests electron withdrawing subsitutents accelerate the reaction (Sello & Di Gennaro, 2013), as expected since the withdrawing group increases electrophilicity at the benzyl carbon, which is the target for nucleophilic attack by the pyruvate enamine. More than one methylene between the aldehyde and aromatic ring such as 3-phenylpropionaldehyde (**Figure 1-5B**) slows down the reaction, unless the extra carbon is part of an α,β-unsaturated system, such as the C-2 carbon in cinnamaldehyde (**Figure 1-5C**). The hydroxyl group in salicylaldehyde (**Figure 1-5A**) increased the rate 2-fold over that of benzaldehyde (Ferrara et al., 2011). The observations suggest that NahE tolerates a wide variety of substrates in the reverse direction. Additionally, the preference for the *o*-hydroxyl group is observed for both the forward and reverse directions.

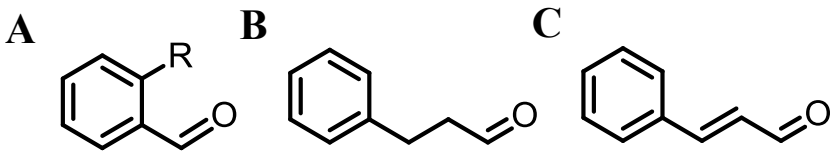
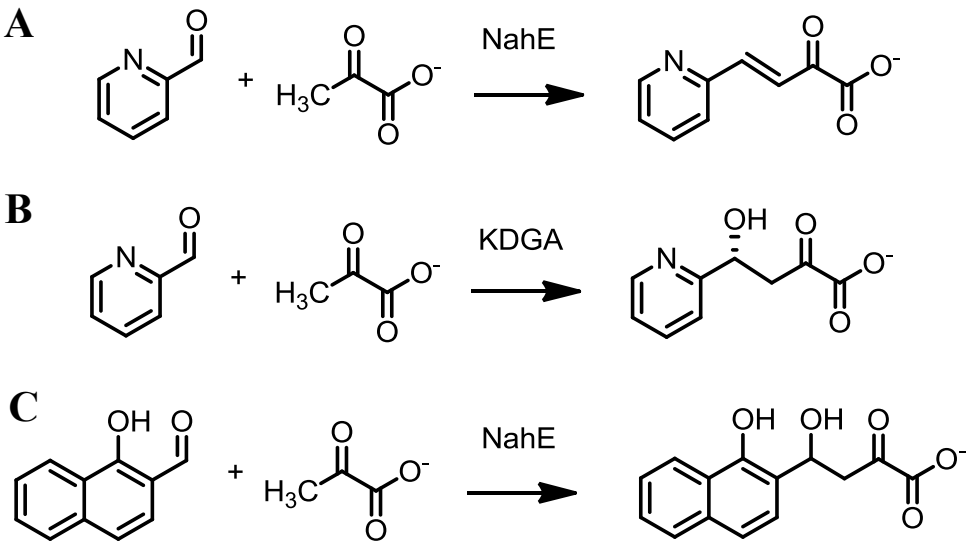


Figure 1-5 Structures of **A)** salicyaldehyde: R=OH, benzaldehyde: R=H **B)** 3-phenylpropionaldehyde **C)** cinnamaldehyde.

The extent to which NahE participates in the dehydration step of the condensation reaction is covered in the literature (Eaton, 2000). This contributes to our understanding of the reaction in the forward direction. The reactions of 3-hydroxybenzaldehyde and 4-hydroxybenzaldehyde with pyruvate give products that have *trans* coupling constant by

^{1}H NMR spectroscopy (Eaton, 2000). This observation indicates that dehydration is enzyme-catalyzed. However, there is the possibility that dehydration occurs quickly so that an intermediate does not accumulate. This possibility could be ruled out by the results of an experiment following the condensation of pyridine-2-carboxaldehyde and pyruvate. When catalyzed by NahE, an olefin is produced (**Scheme 1-6A**), whereas 2-keto-3-deoxy-6-phosphogalactonate aldolase (**KDGA**, **Scheme 1-6B,** S**ection 1.4.5**) produces (*R*)-4-hyrdoxy-2-keto-4-(2′-pyridyl)butyrate (Henderson *et al.*, 1998). In another experiment, mass spectral results show that 1-hydroxy-2-naphthaldehyde, a phenanthrene intermediate, is processed by NahE to form a condensation product that is not dehydrated (**Scheme 1-6C**). The sum of these results suggests the 4-hydroxy intermediate in the reaction with the pyridine and naphthyl aldehyde is stable in solution, and implies dehydration is enzyme-catalyzed when using other aldehydes as substrates.



Scheme 1-6 Evidence for stability of 4-hydroxy intermediates and NahE-catalyzed dehydration in condensation reactions.

Reactions of pyridine-2-carboxaldehyde and pyruvate catalyzed by **A)** NahE and **B)** KDGA, revealing dehydration is catalyzed with NahE, but not with KDGA. **C)** The reaction of 1-hydroxy-2-naphthaldehyde with pyruvate suggests dehydration is not obligatory in NahE catalysis, and that the intermediate is stable.

1.3 PAH METABOLIC NETWORK IN *MYCOBACTERIUM VANBAALENII* PYR-1

M. vanbaalenii PYR-1 completely oxidizes more than ten PAHs including the potent carcinogen benzo[*a*]pyrene using enzymes genomically-encoded in large clusters. Degradation of each ring of the PAH occurs in a sequential manner using the same series of reactions (often catalyzed by different enzymes). The enzymes used in this PAH metabolic network are often nonspecific and redundant in function. Phenanthrene degradation is a focus in this dissertation because it is more complex than napthalene

degradation (**Section 1.2**) and the enzymes involved in its degradation are poorly characterized.

1.3.1 PAHs Degraded and Gene Clustering

M. vanbaalenii PYR-1 is a Gram positive non-pathogenic bacteria that degrades a variety of compounds with varying efficiencies (Heitkamp *et al.*, 1988). The strain was initially isolated from Redfish Bay near Port Aransas, Texas and found to mineralize pyrene (Heitkamp & Cerniglia, 1988, Heitkamp *et al.*, 1988). In addition, *M. vanbaalenii* PYR-1 degraded naphthalene, 2-methylnaphthalene, biphenyl, anthracene, phenenanthrene, fluoranthene, 3-methylcholanthrene, benzo[*a*]pyrene, 1-nitropyrene, and 6-nitrochrysene (**Figure 1-6**, Heitkamp *et al.*, 1988, Heitkamp & Cerniglia, 1989, Kelley & Cerniglia, 1995, Moody *et al.*, 2002, Kweon *et al.*, 2011). When grown in a mixture of PAHs, *M. vanbaalenii* PYR-1 preferred phenanthrene (Kelley & Cerniglia, 1995, MacLeod & Daugulis, 2003).

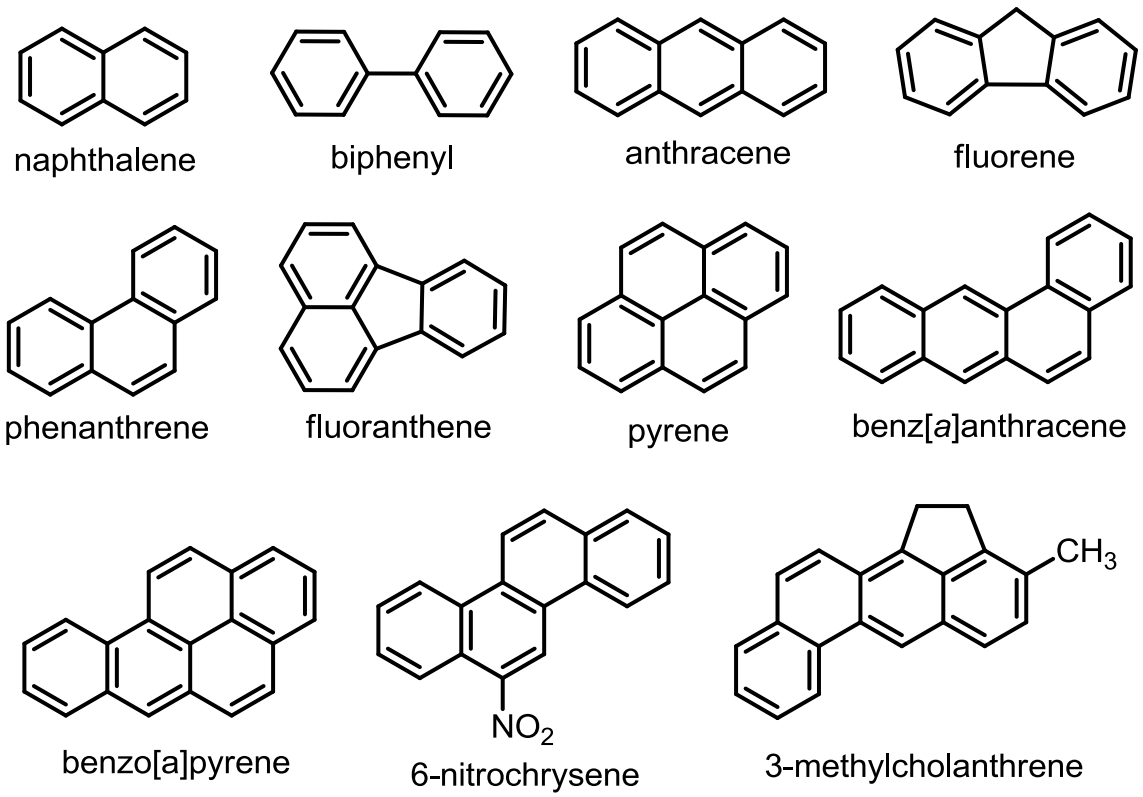


Figure 1-6 PAHs mineralized by *Mycobacterium vanbaalenii* PYR-1.

The circular genome of *M. vanbaalenii* PYR-1 is 6.5 Mb and contains a unique arrangement of genes within two large clusters of 150 kb and 31 kb (Habe & Omori, 2003, Kim *et al.*, 2008). PAH-degrading genes are scattered throughout these clusters in an atypical mosaic structure (where groups of genes that have different evolutionary origins). Just over half of the genes, found in three genomic islands in the 150 kb cluster, were found to have high sequence similarity with nocardioform actinomycetes like *Terrobacter* and *Nocarioides* (Kweon *et al.*, 2011). The 31 kb region is separated from the first by over 2 Mb (Kim *et al.*, 2008). The pattern of gene gain and loss when 27

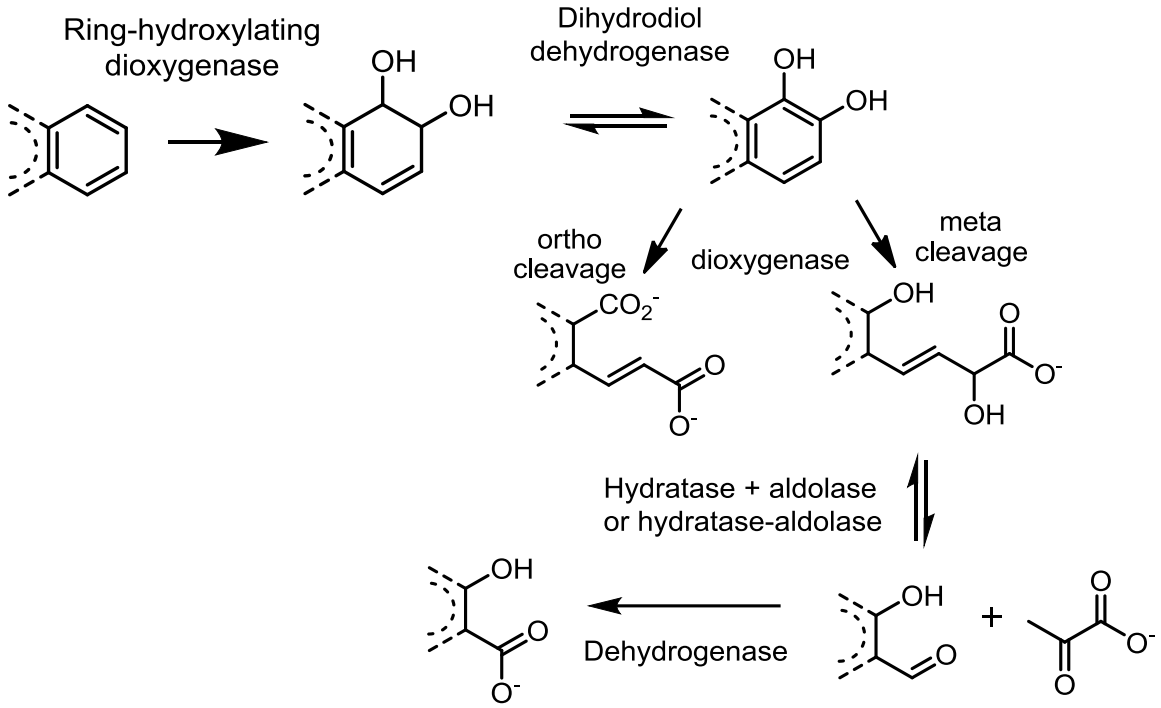
Mycobacterium genomes were compared suggested that PAH degradative ability was acquired from a common ancestor and was dispensable as genus members became pathogenic and host-dependent (Kweon *et al.*, 2015). *Mycobacterium vanbaalenii* PYR-1 is related to *Mycobacterium tuberculosis*, which killed 1.2 million HIV-negative people worldwide in 2010 (WHO, 2011), mostly in developing countries.

1.3.2 Enzymes of the PAH Metabolic Network

Breakdown of the individual rings of a PAH by *M. vanbaalenii* PYR-1 is accomplished sequentially by gene products that are mostly uncharacterized. Each cycle of ring-cleavage and side chain removal (**Scheme 1-7**) is reminiscent of the upper pathway for naphthalene degradation (**Section 1.2**). In fact, *M. vanbaalenii* PYR-1 degrades naphthalene to salicylate, in a pathway that is identical to that in *Pseudomonas putida* G7 (Kelley *et al.*, 1990), except the β-ketoadipate pathway involving ortho cleavage is used in place of the meta fission pathway. Although many enzymes from *Mycobacterium vanbaalenii* PYR-1 have not been characterized, their functions have been inferred from sequence analysis.

The degradation of each individual ring is performed by a collection of diverse enzymes (**Scheme 1-7**, adapted from Kweon *et al.*, 2011). Ring-hydroxylating oxygenases (**RHO**s) are nonheme iron enzymes composed of α and β subunits, and require the electron transfer components ferredoxin and ferredoxin reductase. They are members of the Rieske oxygenase family and use oxygen and NADH to form a *cis*-dihydrodiol (Kweon *et al.*, 2010). Next, Dihydrodiol dehydrogenases use NAD^+ to

oxidize dihydodiols to form dihydroxyl PAHs. Degradation rounds can diverge at this juncture, where ring-cleavage dioxygenases may facilitate *ortho* or *meta* cleavage of PAHs which have *o*-hydroxyl or carboxylate groups. After ring cleavage, the hydratase-aldolases act on the intermediate to remove the side chain. The aldehyde on the aromatic ring is often oxidized to a carboxylic acid by an aldehyde dehydrogenase using NAD^+. A decarboxylase might act on the aromatic acid before the next cycle begins.



Scheme 1-7 General cycle for the degradation of a single ring of a PAH.

Central metabolic processes such as glycolysis and the Krebs cycle are important for integrating metabolic networks. The same concept is found in *M. vanbaalenii* PYR-1, because all degradative pathways converge at the monocyclic metabolite phthalate, the

hub of the metabolic network. Degradation products proceed through β-ketoadipate pathway (Kweon *et al.*, 2011, Stingley *et al.*, 2004), where *ortho* cleavage occurs and the final products are succinyl-CoA and acetyl-CoA. The central aromatic process is loosely regulated in comparison to the tightly regulated upper pathways (Kweon *et al.*, 2014). The funneling of PAH metabolites into one single pathway (the β-ketoadipate pathway) for monocyclic aromatic hydrocarbon degradation simplifies the network and regulation system.

In the course of these reactions, *o*-quinones are generated by oxidation of PAH catechols. These quinones modify cell components and generate reactive oxygen species. *M. vanbaalenii* PYR-1 detoxifies *o*-quinones by converting them to catechols using quinone reductase and NAD(P)H (Kim *et al.*, 2003). Methyltraneferases modify Non-K and non-Bay PAH catechols (**Figure 1-7**) of anthracene, fluoranthene, and benzo[*a*]pyrene for excretion (Kelley *et al.*, 1993, Moody *et al.*, 2001, Moody *et al.*, 2004).

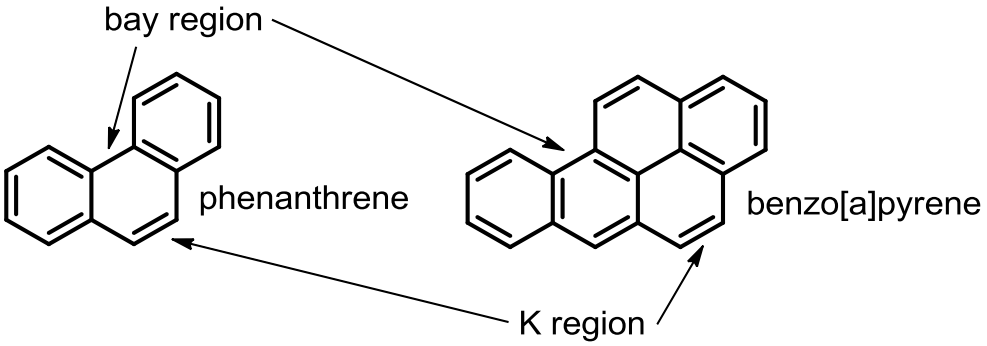


Figure 1-7 Bay and K regions of several PAHs.

Though *o*-quinone PAHs are often biologically conjugated by glutathione (Bolton *et al.*, 2000), *M. vanbaalenii* PYR-1 might uses mycothiol as a substitute for glutathione. This subsititution is evident in *M. tuberculosis*, where mycothiol biosynthesis is obligatory (Sareen, *et al.*, 2003) and to which glutathione is directly toxic (Dayaram, *et al.*, 2006). A glutathione *S*-transferase enzyme (Mvan_5043) was identified in *M. vanbaalenii* PYR-1 and proposed to be involved in detoxification (Kim *et al.*, 2008). However, the protein could be an isomerase, like the glutathione *S*-transferase NahD (18.8% sequence ID) which participates in naphthalene degradation (**Section 1.2**). The transferase and isomerase activity both require glutathione, and it is unknown if mycothiol could substitute for the substrate/cofactor in Mvan_5043 from *M. vanbaalenii* PYR-1.

1.3.3 Redundancy and Pleiotropy of Enzymes in *M. vanbaalenii* PYR-1

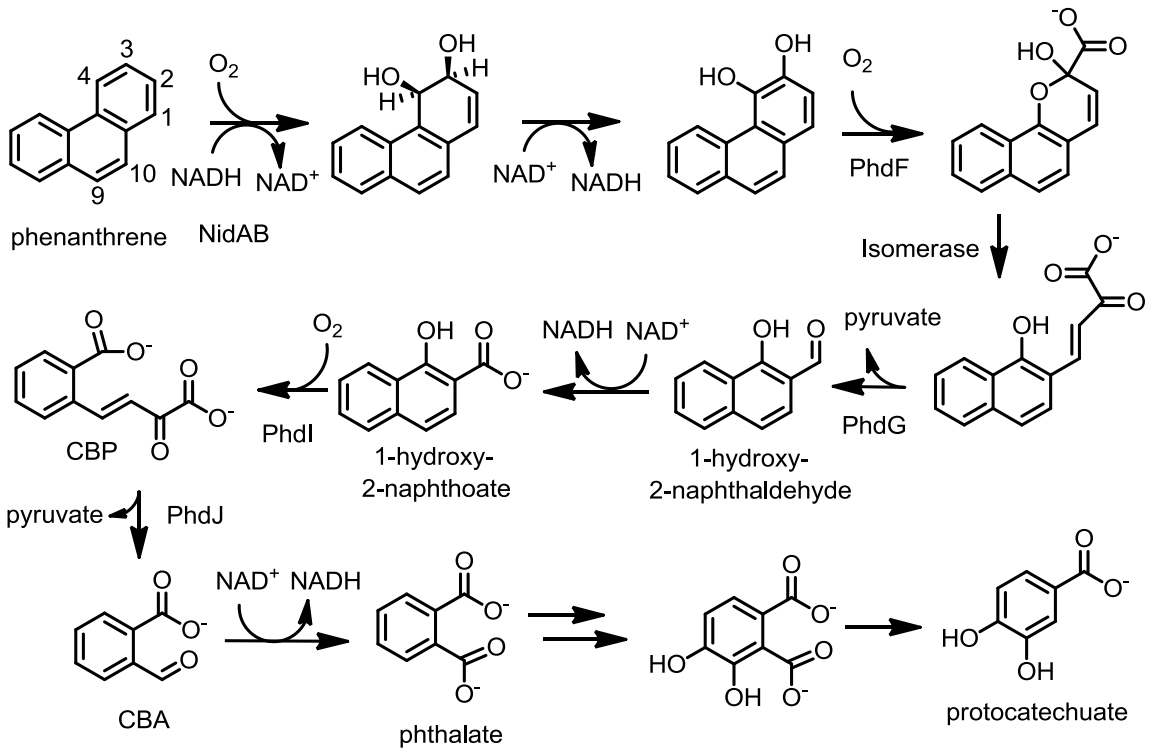
The number of each type of enzyme in the PAH degradative network in *M. vanbaalenii* PYR-1 varies. When the *M. vanbaalenii* PYR-1 proteome was analyzed in after exposure to 7 aromatic compounds (Kweon *et al.*, 2011), commonly upregulated proteins included 10 ring-hydroxylating oxygenases (RHOs), 20 cytochrome P450s (CYPs), 10 other monooxygenases, two epoxide hydrolases, two dihydrodiol dehydrogenases, five ring-cleavage dioxygenases, five hydratase-aldolases, five hydrolases, and only one decarboxylase.

Many RHOs in *M. vanbaalenii* PYR-1 have overlapping specificities. For example, *in vivo* experiments with recombinant PdoA2B2, NidA3B3, and NidAB show that each can oxidize phenanthrene at the 3,4 position, at varying efficiencies (Kweon *et al.*, 2014). There are a number of reasons *M. vanbaalenii* PYR-1 might produce redundant enzymes for a particular activity. Redundancy may exist so that a particular activity can be modulated incrementally by varying expression of isozymes with different kinetic parameters, so that expression is not a binary on/off switch. This effect would also prevent a lag while the necessary enzyme is produced. Redundancy could also exist as an evolutionary strategy to increase the number of compounds degraded by *M. vanbaalenii* PYR-1, by "tweaking" active sites to accommodate new substrates.

In addition to the redundancy of enzymes that catalyze a single reaction, many enzymes from *M. vanbaalenii* PYR-1 are nonspecific, which is considered pleiotropy. For example, NidA3B3 was shown to process six different PAHs, including both LMW and HMW species (Kim *et al.*, 2006). Pleiotropy is also observed in the RHO components; all RHOs in *M. vanbaalenii* PYR-1 share the same ferredoxin and ferredoxin reductase enzymes encoded by *phtAcAd* (Kim *et al.*, 2007). Furthermore, a single decarboxylase was detected when *M. vanbaalenii* PYR-1 was incubated with various PAHs. Although it seems wasteful for *M. vanbaalenii* PYR-1 to express multiple nonspecific enzymes, some metabolites may only be processed by one nonspecific enzyme, and the ability of that particular enzyme to catalyze other reactions may not be harmful for the bacteria.

1.3.4 Phenanthrene Catabolism in *M. vanbaalenii* PYR-1

In the presence of phenanthrene, heterologous expression of a 37-kb region (cloned from the larger 150 kb cluster of PAH catabolism genes into a fosmid) of the *M. vanbaalenii* PYR-1 genome in the presence of phenanthrene generated phenanthrene *cis*-3,4-dihydrodiol. In the presence of 1-hydroxy-2-naphthoic acid, the culture produced *trans*-*o*-carboxybenzylidenepyruvate (**CBP**), *o*-carboxybenzaldehyde (**CBA**), and phthalic acid (Stingley *et al.*, 2004). A catabolic pathway was proposed (**Scheme 1-8**) using enzymes encoded in the fosmid (including those shown in **Figure 1-8**). Other data show that addition to hydroxylation at the 3 and 4 positions, phenanthrene is oxidized to *cis*-dihydrodiols at the 1,2- and 9,10- positions, the latter of which is an alternative degradation route (Kim *et al.*, 2005). The numbering system for phenanthrene is shown in **Scheme 1-8**.



Scheme 1-8 Phenanthrene catabolic pathway via 3,4-dihydroxyphenanthrene.

Stingley *et al.* propose the 3,4-dihydroxyphenanthrene ring-cleavage product is processed by an isomerase that is missing from the cloned 37-kb region to form a 3,4-unsaturated α-keto acid that is converted by the *phdG* gene product, annotated as a hydratase-aldolase. Pyruvate and 1-hydroxy-2-naphthaldehyde are produced, and the latter is then oxidized by an NAD^+-dependent aldehyde dehydrogenase to form 1-hydroxy-2-naphthoate. *ortho* cleavage of this compound by PhdI forms CBP. The *cis* (Z) isomer of CBP is the reasonable first product of the dioxygenase, but deuterium exchange at C-3 did not occur in D_2O during the reaction with PhdI. Thus, the *cis* to *trans*

conversion was proposed to be part of the catalytic mechanism (Adachi *et al.*, 1999). The ring-cleavage product CBP is then processed by *trans*-*o*-carboxybenzylidenepyruvate hydratase-aldolase (**PhdJ**), to form CBA and pyruvate. An aldehyde dehydrogenase converts CBA to phthalate, which is degraded via the β-ketoadipate pathway and the Krebs cycle.

1.3.5 The Hydratase-Aldolases PhdJ and PhdG

Five hydratase-aldolases are proposed to be involved in PAH metabolism and vary significantly in terms of expression levels in the presence of 7 different PAHs (Kweon *et al.*, 2011). Hydratase-aldolase genes *phdJ* and *phdG*, proposed to be involved in phenanthrene catabolism (**Section 1.3.4**), are located in the same operon (**Figure 1-8**, Stingley *et al.*, 2004). The presence of pyrene increased the expression of PhdJ and PhdG about two-fold. Although both enzymes were not upregulated when *M. vanbaalenii* PYR-1 was incubated with fluoranthene, the expression levels of putative hydratase-aldolases Mvan_0558 and Mvan_2387 increased (Kweon *et al.*, 2007, Kim *et al.*, 2008).

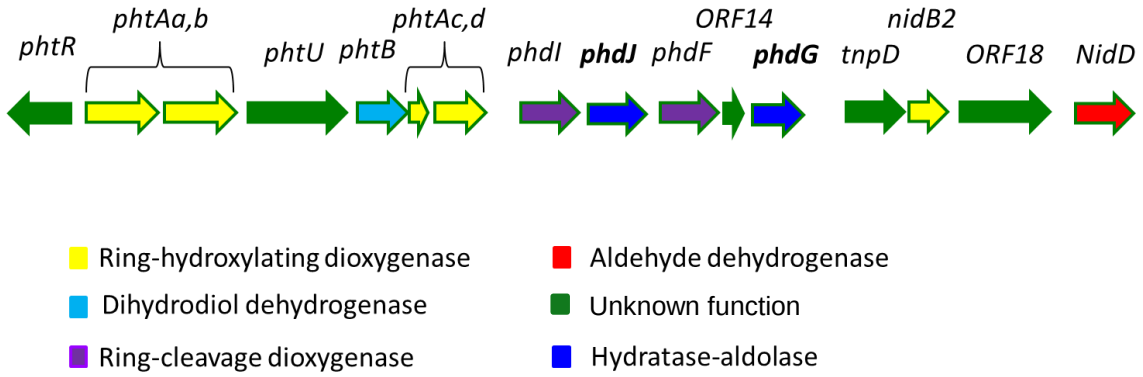


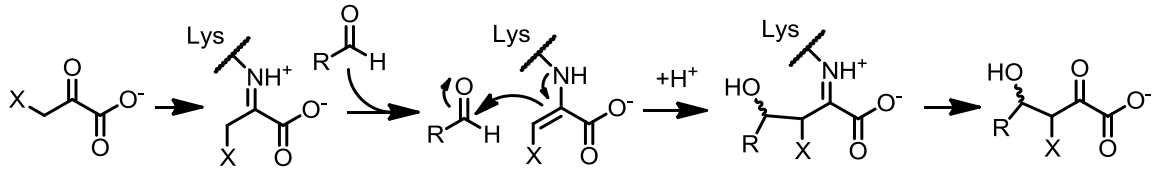
Figure 1-8 Genomic context of PhdJ and PhdG.

Sequence analysis identified PhdJ as a *N*-acetylneuraminate lyase (NAL) subgroup member in the Class I aldolase superfamily (**Section 1.4**). PhdJ from *Nocardioides* KP7 was isolated and shown to function as a hydratase-aldolase (Iwabuchi & Harayama, 1998). Conversion of CBP to CBA and pyruvate by purified PhdJ (see **Scheme 1-8** for structures) was ~95% complete at equilibrium. The k_{cat} and K_M values were 13 s^{-1} and 50 μM (k_{cat}/K_M = 2.6 × 10^5 $M^{-1}s^{-1}$), respectively. The enzyme doesn't accept *trans*-*o*-hydroxycinnamate, BP, or *trans*-*o*-methoxybenzylidenepyruvate (**Figure 1-4**) as substrates. Surprisingly, the incubation of PhdJ with pyruvate (20 μM) and $NaBH_4$ did not inhibit the enzyme, as would be expected for type I aldolases (see the next section).

1.4 THE *N*-ACETYLNEURAMINATE LYASE SUB-FAMILY OF TYPE I ALDOLASES

1.4.1 Type I Aldolases

Class I aldolase superfamily members share the $(\alpha/\beta)_8$, or triose phosphate isomerase (TIM) barrel motif and catalyze aldol reactions using α-keto acids as electron donors and aldehydes as electron acceptors via Schiff base intermediates (**Scheme 1-9**). Schiff base formation and hydrolysis with a conserved lysine and substrate is facilitated by a conserved glutamate, tyrosine, or histidine on β-strand 5 next to the conserved lysine on β-strand 6 (Choi *et al.*, 2006). The superfamily includes the glycolytic enzyme aldolase, transaldolase, and members of the *N*-acetylneuraminate lyase (NAL) subgroup.



Scheme 1-9 Aldol mechanism for Class I Aldolases.

1.4.2 The *N*-Acetylneuraminate Lyase (NAL) Subgroup

The NAL subgroup of type I aldolase enzymes participate in diverse reactions including those in PAH catabolism (**Sections 1.4.3-1.4.6**). Like all type I aldolases, members of the NAL subgroup have a conserved lysine residue that participates in covalent catalysis. During reactions catalyzed by enzymes of the NAL subgroup, a Schiff

base is generated using the ε-amine group of a conserved lysine on β-strand 6 and the 2-oxo group of substrate. A conserved tyrosine is on the adjacent strand, β-strand 5, and is the general acid/base that mediates Schiff base formation. Enzymes in the NAL subgroup are identified by their inactivation by pyruvate and other α-keto acid substrates in the presence of sodium borohydride or sodium cyanoborohydride. This reduces Schiff base moiety to produce an amino acid, and tethers the substrate or ligand to the conserved lysine.

Enzymes in the NAL subgroup are tetrameric, where each subunit is composed of an $(\alpha/\beta)_8$ barrel and three C-terminal α-helices (**Figure 1-9A,C**). Members of the NAL subroup have "tight" and "weak" interfaces between monomers. Most NAL subgroup members orient the three C-terminal helices near the weak interface (parallel to the sheets of the β-barrel, **Figure 1-9A**), with the exception of plant DHDPS, where the helices point in the direction normal to the plane of the tetramer (Atkinson *et al.*, 2012). The tight interface is formed by a tryptophan or tyrosine located in the loop between β-strand 4 and α-helix 4 that extends deep into an adjacent subunit. In the available structures, the hydrophobic side chain of Trp or Tyr breaches various loops near the N-terminal region of the β-barrel and forms a hydrogen bond with the side chain of serine or threonine in the neighboring subunit positioned as the second X in a conserved GXXGE motif. The motif, arranged in two overlapping type I β-turns, is placed between β-strand 2 and α-helix 2 (**Figure 1-9B,C**), where the backbone amides interact via hydrogen bonds with the C-1 carboxylate group of an imine adduct of pyruvate (**Figure 1-9D**), for example.

The conserved lysine, tyrosine (at β-strand 5), and GXXGE motif constitute the primary residues that are conserved throughout the NAL subgroup.

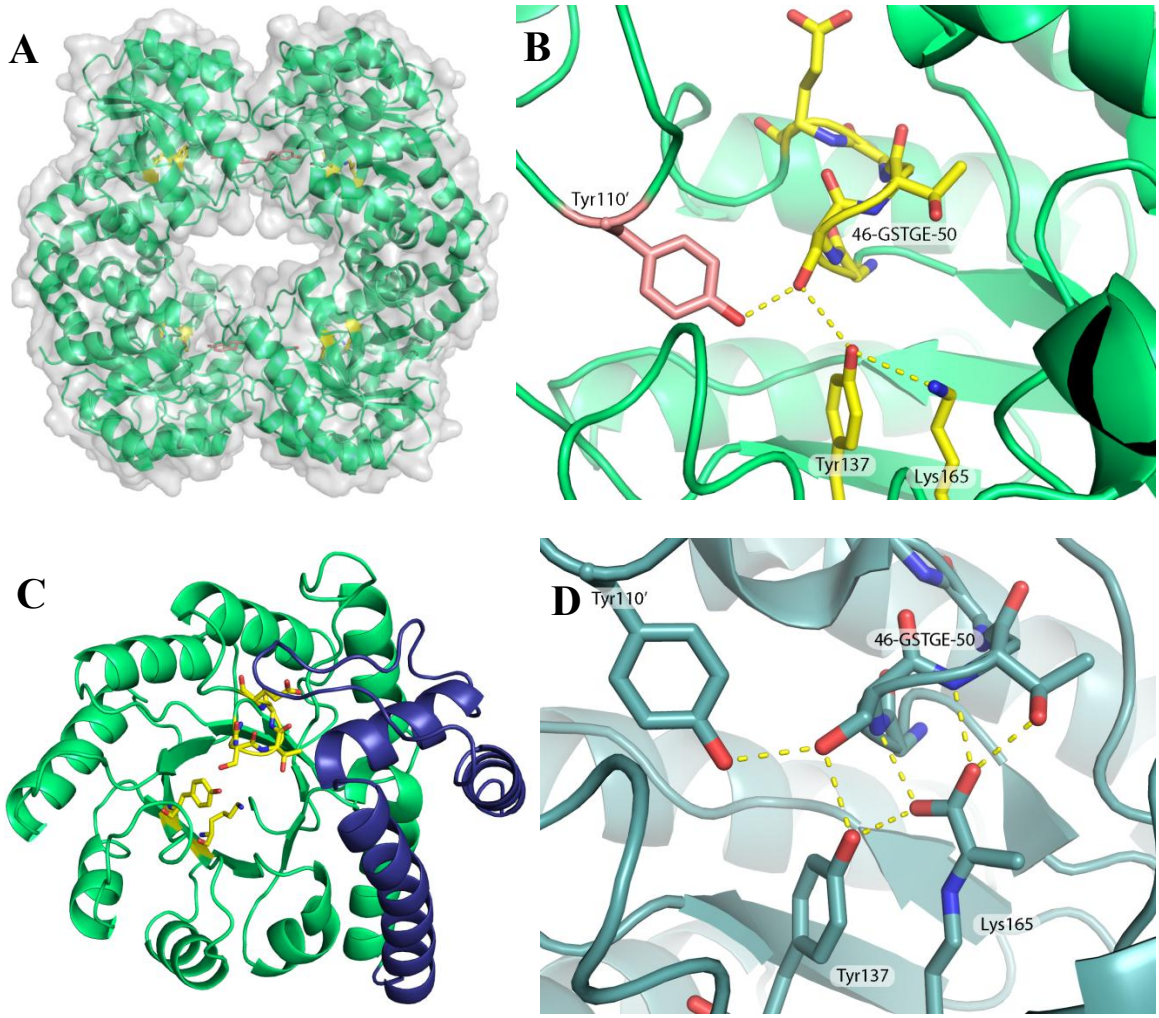


Figure 1-9 The crystal structure of *E. coli* NAL (2.2 Å resolution, PDB entry 1NAL). **A)** Tetrameric structure with surface representation. **B)** Intersubunit interaction in the tight interface showing Tyr-137, Lys-165, GXXGE motif (yellow) and Tyr110′ (salmon). **C)** $(\alpha/\beta)_8$ of NAL (lime green) showing three C-terminal helices (dark blue). **D)** Active site of *E. coli* NAL (light teal) where Lys-165 is covalently attached to pyruvate (1.7 Å resolution, PDB entry 2WNN).

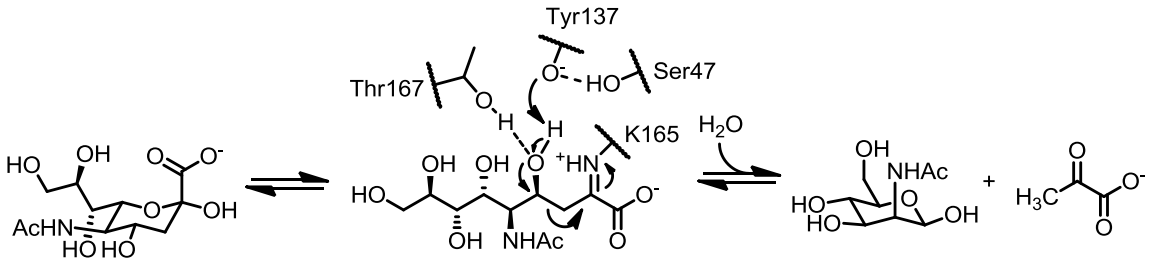
1.4.3 *N*-Acetylneuraminate Lyase

N-Acetylneuramintate lyase (NAL) converts the 9 carbon sugar *N*-acetylneuraminic acid, or sialic acid, to *N*-acetyl-D-mannosamine and pyruvate. Sialic acid is toxic to *E. coli*, and detection of the amino sugar by the bacterium is followed by induction of NAL. The kinetic mechanism is Uni-Bi ordered sequential, where sialic acid binds and is processed, to form acetyl-D-mannosamine and pyruvate. Acetyl-D-mannosamine leaves prior to pyruvate. NAL has the conserved lysine, tyrosine, and GXXGE motif, but the enzyme also has different active site residues closer to the entrance of the active site cavity to accommodate the hydrophilic substrate during the reaction.

The crystal structure of the *E. coli* NAL with the pyruvoyl adduct shows that the C-1 carboxylate of pyruvate interacts with backbone amides of Ser-47 and Thr-48 of the GXXGE motif as well the hydroxyl group of Tyr-137 (**Figure 1-9D**) (Barbosa *et al.*, 2000). The side chains of Tyr-110′ (prime indicates adjacent NAL monomer), Ser-47, and Tyr-137 hydrogen bonds and link the adjacent monomer to the C-1 carboxylate group of pyruvate. Although the interactions of Tyr-110′ and Ser-47 suggest that Tyr-110′ stabilizes the catalytic Tyr-137 hydroxyl group, its hydrophobic property seems to be more important. The Y110A variant shows a 40-fold drop in k_{cat}/K_M and the Y110F variant did not affect activity significantly (Daniels *et al.*, 2014). The bulk and hydrophobicity of this residue is shown to be more important in other NAL subgroup members as well. The conserved lysine, tyrosine, and GXXGE motif are considered the

primary residues that are conserved in all NAL subgroup members, and all other residues that contribute to binding and catalysis are specific to individual enzymes in the subgroup and are designated secondary residues.

Crystallographic and kinetic experiments suggest that a proton is transferred from the hydroxyl group at C-4 of sialic acid to Tyr-137 (**Scheme 1-10**), a residue shown to be critical for activity, with the assistance of the hydroxyl group of Thr-167 (Barbosa *et al.*, 2000, Daniels *et al.*, 2014). Thr-167 is not conserved in DHDPS, KDGDH, NahE, or PhdJ, but is present in KDGA. QM/MM simulations suggest deprotonation is rate-determining as the activation barrier is mostly due to C-C bond formation (Daniels *et al.*, 2014). Modeling of Schiff base intermediates in the active site of NAL indicated direct substrate-assisted catalysis where Tyr-137 is deprotonated by the C-1 carboxylate of the pyruvoyl adduct in preparation for retro-aldol cleavage (Smith *et al.*, 1999). However, QM/MM simulations suggest that a Tyr-137 phenoxide is stable and does not need to be protonated in a substrate-assisted manner (Daniels *et al.*, 2014).



Scheme 1-10 Mechanism for NAL-catalyzed conversion of N-acetyneuraminate (sialic acid) to acetyl-D-mannosamine and pyruvate.

The backbone carbonyl of Gly-188 may also contribute to catalysis. In a structure of the *Haemophilus influenza* homologue, Gly-188 is closer than Thr-166 (*H. influenza* numbering) to a C-4 *gem*-diol of a sialic acid analogue bound to NAL (**Figure 1-10,** all residues shown are conserved in the *E. coli* enzyme, Barbosa *et al.*, 2000). This interaction might allow stabilization of the partial negative charge on the C-4 oxygen of substrate that develops during deprotonation.

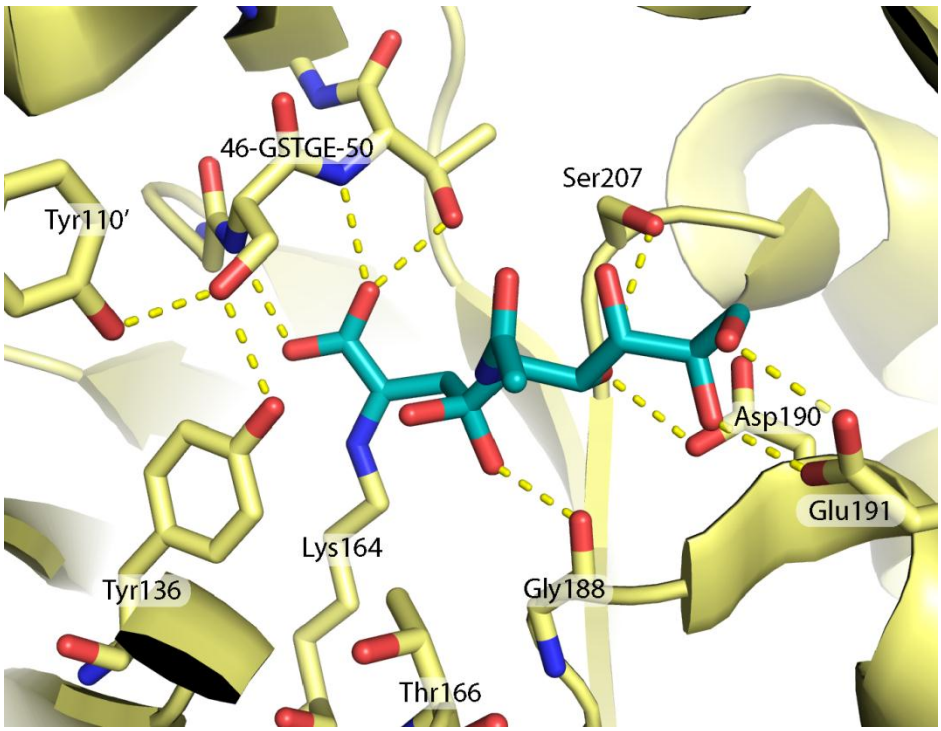
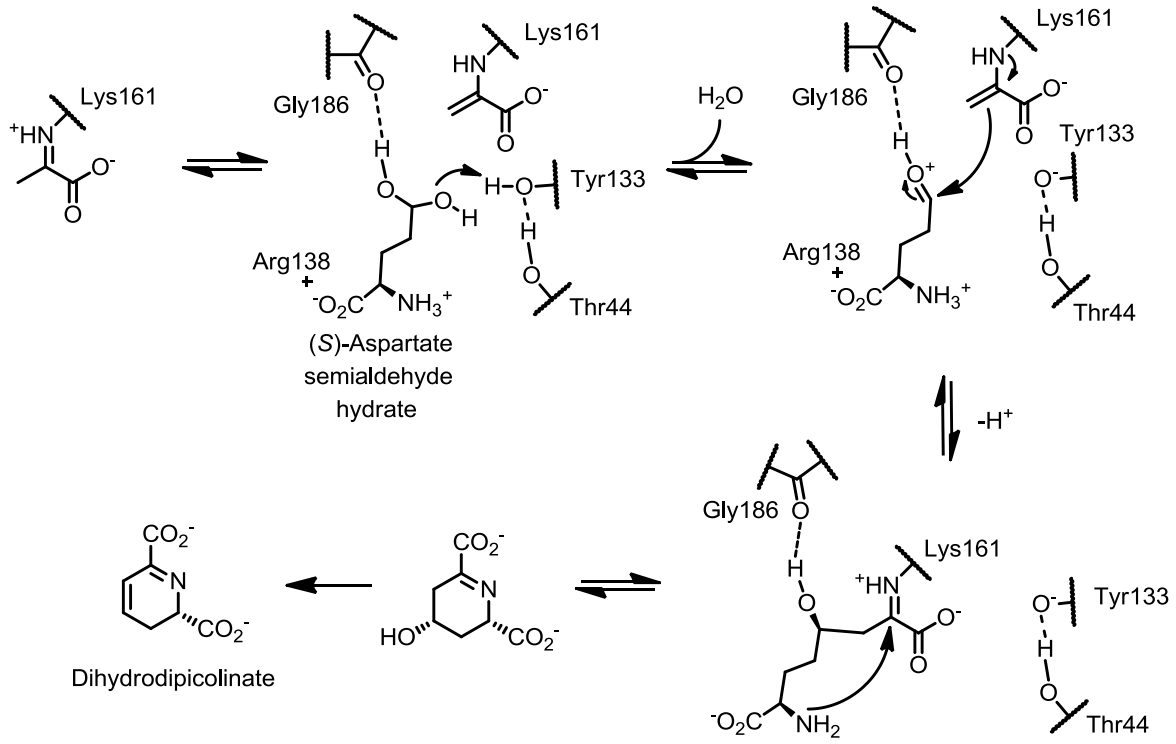


Figure 1-10 The crystal structure of *Haemophilus influenza* NAL (1.6 Å resolution, PDB entry 1F7B) liganded with the C-4 *gem*-diol of substrate analogue *N*-acetyl-4-oxoneuraminate.

1.4.4 Dihydrodipicolinate Synthase

Dihydrodipicolinate synthase (DHDPS) catalyzes the condensation of the *gem*-diol form of (*S*)-aspartate-β-semialdehyde and pyruvate to form (*S*)-4-hydroxy-2,3,4,5-tetrahydro-(2*S*)-dipicolinic acid and participates in the *meso*-diaminopimelate pathway for lysine biosynthesis (**Scheme 1-11,** Blickling *et al.,* 1997, Dobson *et al.,* 2008, Dobson *et al.*, 2009). After the pyruvoyl Schiff base is formed, presumably with the assistance of Tyr-133, tautomerization allows the formation of an enamine. After binding of (*S*)-aspartate semialdehyde hydrate (the reaction is Bi-Uni), dehydration occurs using Tyr-133, while the geminal hydroxyl group interacts with the backbone carbonyl of Gly-186. The protonated aldehyde is susceptible to attack by C-3 of the enamine adduct. The aldol reaction gives a 7-carbon adduct, which undergoes cyclization via attack of the α-amine. Rearrangement to form the imine in the ring leads to elimination of Lys-161. After the product, (*S*)-4-hydroxy-2,3,4,5-tetrahydro-(2*S*)-dipicoolinic acid, leaves the active site, dehydration occurs and dihydrodipicolinate is formed.



Scheme 1-11 DHDPS-catalyzed condensation of pyruvate and (*S*)-aspartate semialdehyde hydrate.

Like NAL, a triad of residues in DHDPS links the catalytic tyrosine (Tyr-133) to the GXXGE motif in the neighboring subunit. For the *E. coli* homologue, these residues are Tyr-133, Thr-44, and Tyr-107′ (**Figure 2-11**). As for NAL, the second residue of the GXXGE motif in DHDPS is critical (Laber *et al.*, 1992), and the hydroxyl group of Tyr-107′ is less important. However, in contrast with NAL, the hydroxyl group of Tyr-107′ in DHDPS is significant as the relative activity of the Y107F mutant is 17% of the wild-type activity. The decrease indicates that the hydroxyl group plays a role in catalysis, as opposed to solely strengthening the tight interface through hydrophobic intreactions, as

shown for NAL (Griffin *et al.*). The size of Tyr-107′ preserves the tetrameric state and activity of DHDPS. The Y107W mutant is mostly monomer and tetramer and the k_{cat} decreases 20-fold (Pearce *et al.*, 2008). Tyr-107 is necessary for tetramerization and may protect the active site from solvent or provide structural stability that is critical for activity.

While the carboxylate of pyruvate is stabilized by the GXXGE motif, Arg-138 forms an important electrostatic interaction with the carboxylate of (*S*)-aspartate-β-semialdehyde hydrate. This residue is critical for catalysis, as a more than 100-fold reduction in k_{cat} was observed for R138A and R138H variants relative to the wild-type activity (Dobson *et al.* 2005).

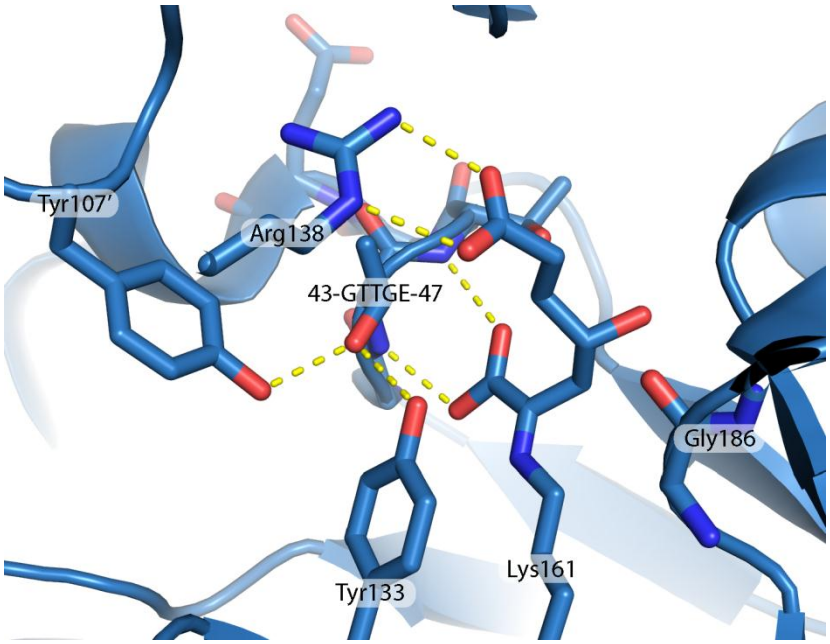
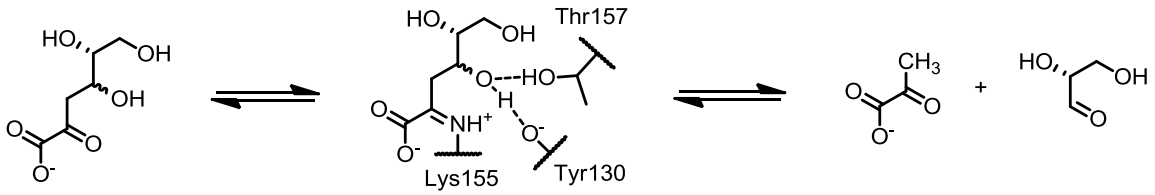


Figure 1-11 The crystal structure of the *E. coli* dihydrodipicolinate synthase covalently bound with condensation product of pyruvate and succinate (2.3 Å resolution, PDB entry 4EOU).

1.4.5 D-2-Keto-3-deoxygluconate Aldolase

D-2-Keto-3-deoxygluconate Aldolase (KDGA) from the thermophilic Archaea *Sulfolobus solfataricus* has a half-life of 2.5 hours at 100°C and participates in a variant of the Entner-Doudoroff glucose catabolic pathway. The enzyme is promiscuous, and can convert either D-2-keto-3-deoxygluconate or D-2-keto-3-deoxygalactonate to D-glyceraldehyde and pyruvate using Tyr-155, Lys-130, and Thr-157 as catalytic residues (Theodossis *et al.*, 2004, **Scheme 1-12**).



Scheme 1-12 The KDGA-catalyzed conversion of D-2-keto-3-deoxygluconate or D-2-keto-3-deoxygalactonate to D-glyceraldehyde and pyruvate.

Crystal structures reveal space to accommodate and stabilize the C-5 and C-6 hydroxyl groups of substrate at different positions in the active sites, explaining the lack of specificity for KDGA (Theodossis et al., 2004). The backbone carbonyl of Gly-179 participates in C-C bond cleavage. It is conserved in NAL and DHDPS, but its role in assisting Tyr-130 (KDGA numbering) in the deprotonation of the C-4 hydroxyl group of its substrate is likely substituted by Thr-157.

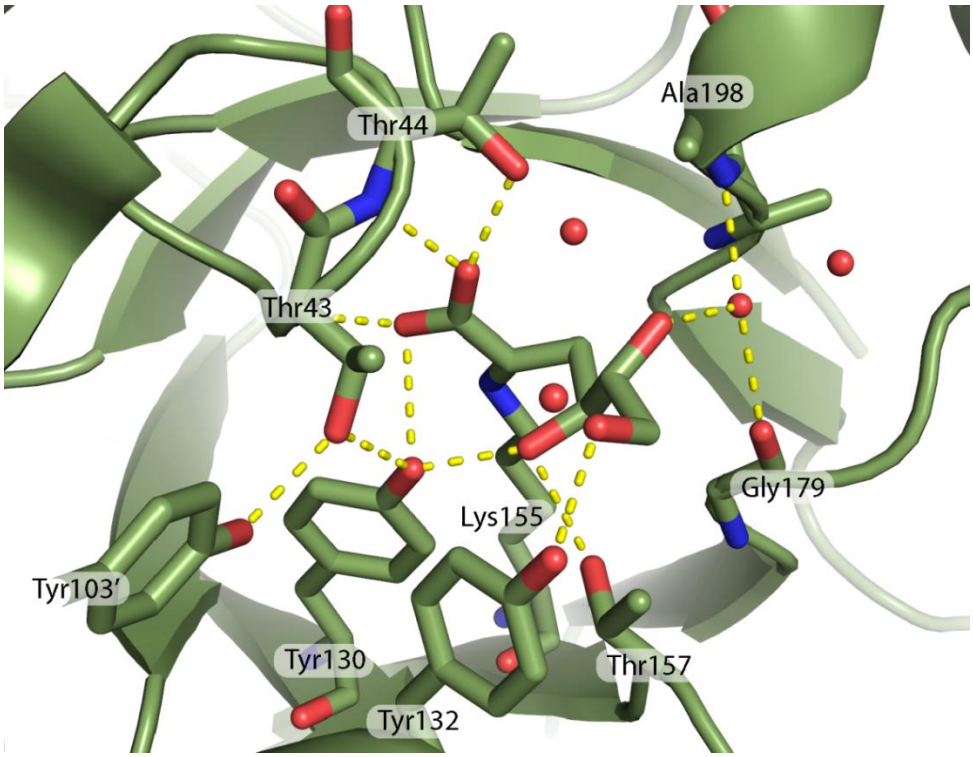


Figure 1-12 Active site of the KDGA showing D-2-keto-3-deoxygluconate covalently bound at Lys-155 (2.1 Å resolution, PDB entry 1W3N).

1.4.6 Keto-deoxy-D-galactarate Dehydratase

Keto-deoxy-D-galactarate dehydratase (KDGDH) from *Agrobacterium fabrum* catalyzes the dehydration and decarboxylation of keto-deoxy-D-galactarate, or 3-deoxy-L-*threo*-2-hexulosarate, forming α-ketoglutaric semialdehyde and CO_2 (**Scheme 1-13**). The enzyme is a tetramer and participates in the catabolism of D-galactouronate, one of the major components of pectin. Unlike all other subgroup members, the enzyme does not have a serine or threonine at the second X of the GXXGE motif, which interacts with the conserved tyrosine.

The crystal structure of KDGDH labeled with 2-oxoadipate, a 3,4-dideoxy analogue, reveals that Arg-142 replaces Ser/Thr in the GXXGE motif, which stabilizes Tyr-140 (**Figure 1-13**). The tyrosine that typically links Ser/Thr with a neighboring monomer is present at position 113, but is positioned several residues N-terminal to the usual location. The lack of the triad typically present in tetrameric KDGDH is consistent with the findings for DHDPS and NAL where the hydrophobic property of the tyrosine is more important for maintaining the tetrameric state and retaining enzyme activity than is hydrogen bonding using the Tyr-Ser/Thr-Tyr triad.

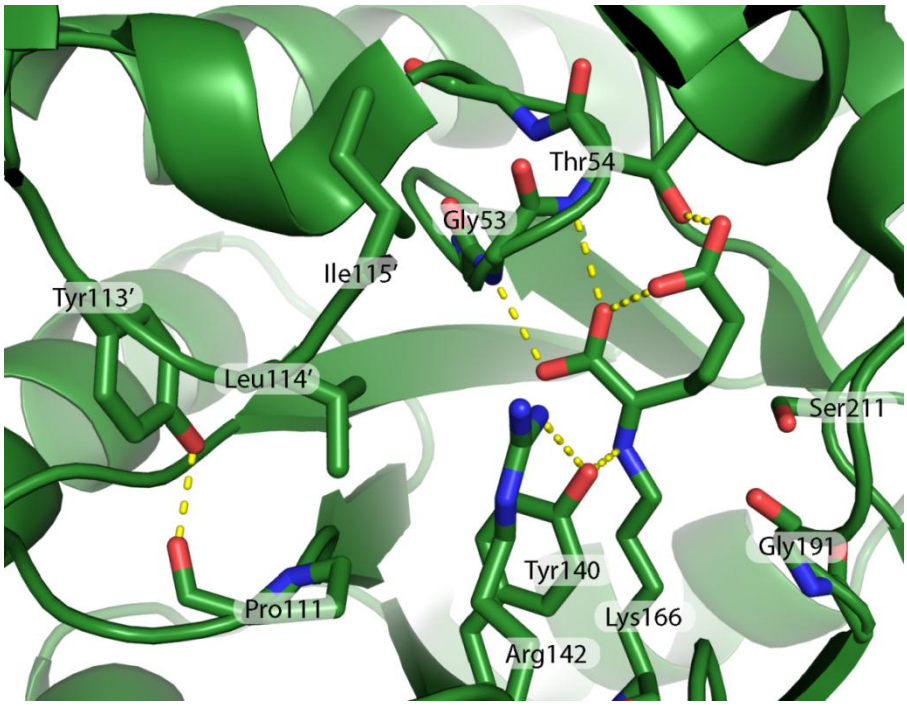
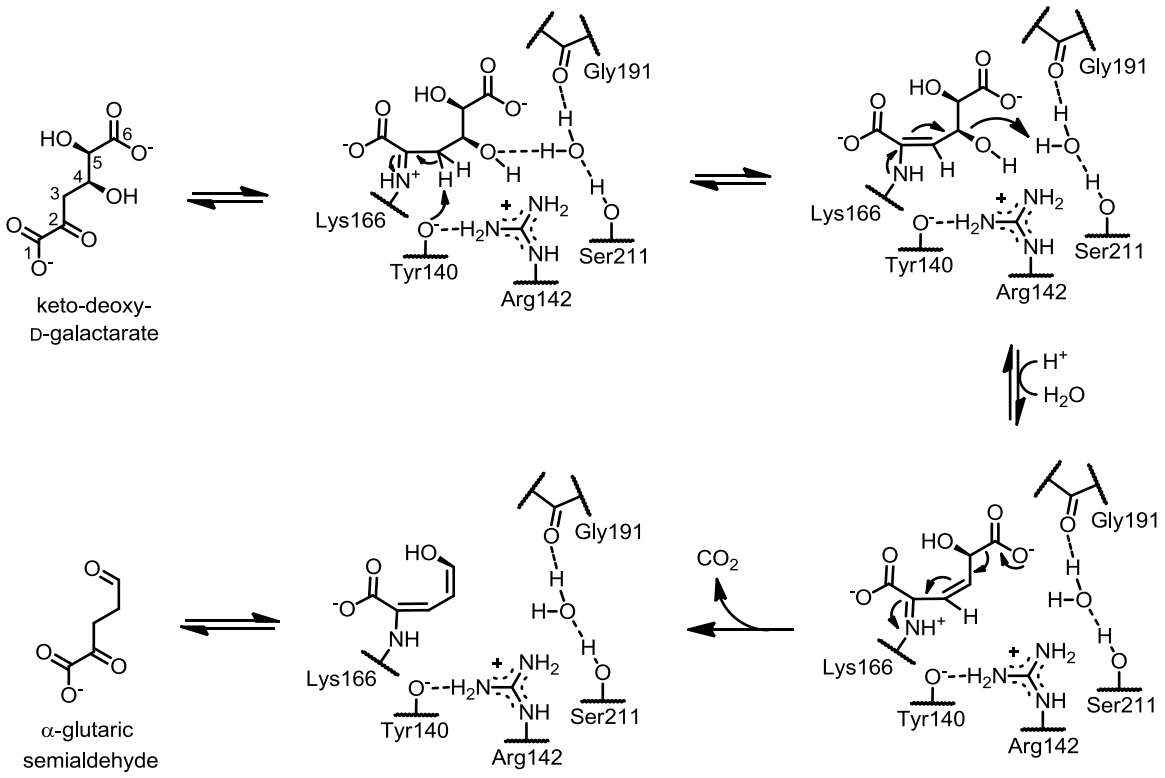


Figure 1-13 Active site of the KDGDH covalentely bound with 2-oxoadipic acid (1.9 Å resolution, PDB entry 1EC8).

The proposed mechanism for decarboxylation of keto-deoxy-D-galactarate is shown in **Scheme 1-13** (Jeffcoat *et al.*, 1969, Taberman *et al.*, 2014). After Schiff base formation with Lys-166 and substrate, dehydration begins with deprotonation at C-3 and formation of an enamine using Tyr-140, which is shown to be nearby C-3 in the crystal structure (**Figure 1-13**). An α,β-unsaturated imine is then formed concurrently with dehydration. Protonation of the leaving hydroxyl group is likely facilitated by the side chain of Ser-211 and backbone carbonyl of Gly-191 (conserved in NAL, DHDPS, and KDGA) via a water molecule, due to proximity in the crystal structure and the 30-fold reduction in specific activity of the S211A variant. The adduct is then poised for decarboxylation, where electrons flow from the anionic carboxylate to the imino nitrogen of Lys-166 with release of CO_2. The resulting dienol tautomerizes twice, forming an aldehyde and iminium. This adduct is then hydrolyzed and released as 2-oxo-glutaric semialdehyde.



Scheme 1-13 Proposed mechanism for KDGDH.

Chapter 2: The Hydratase-Adolase Homologues NahE, PhdJ, and PhdG

2.1 Introduction

Sequence analysis shows that NahE, PhdJ, and PhdG have Lys-183, Tyr-155, and the GXXGE motif, which are conserved in all other NAL subgroup members. Studies of these enzymes suggest that Tyr-155 is involved in Schiff base formation between the lysine and substrate and the subsequent aldol reaction. However, the other residues that contribute to these steps and those that activate water in the hydration are unknown. Additionally, residues that provide selectivity for NahE and PhdJ have not been identified. Kinetic and crystallographic studies allowed us to identify secondary active site residues that are unique to hydratase-aldolases. Based on the positions in liganded and unliganded structures of NahE and PhdJ, we are able to propose roles for each residue in binding and catalysis for these enzymes. The proposed activity for PhdG in the phenanthrene catabolic pathway could not be confirmed in our hands. However, we showed activity with other substrates including HBP, and determined crystal structures. These structures were compared with those of NahE and PhdJ.

2.2 Experimental Procedures

Materials. Chemicals, biochemicals, buffers, and solvents were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO), Fisher Scientific Inc. (Hampton, NH), or EMD Millipore, Inc. (Billerica, MA). The Phenyl-Sepharose 6 Fast Flow, DEAE Fast

Flow, and Superdex 200 resins used for protein purification was obtained from GE Healthcare Life Sciences (Boston, MA). The HisPur Ni-NTA resin was obtained from ThermoFisher Scientific (Waltham, MA). The Econo-Column® chromatography columns were obtained from Bio-Rad (Hercules, CA). The Amicon stirred cell concentrators and the ultrafiltration membranes (10000 and 30000 Da, MW cutoff) were purchased from EMD Millipore Inc. Oligonucleotide primers were obtained from Sigma-Aldrich.

Crystallography. Crystallization experiments were carried out in the Macromolecular Crystallography Facility in the Institute for Cellular and Molecular Biology (ICMB) at the University of Texas at Austin. Various crystallization conditions were screened using the Phoenix crystallization robot made by Art Robbins Instruments (Sunnyvale, CA). Four crystallization screening kits (JCSG+, PEGs, Protein Complex, and Classics Suites) were obtained from Qiagen (Venlo, Netherlands). Each kit contains 96 conditions that were screened in three ratios of protein to reservoir solution volume (1:2, 1:1, and 2:1). Conditions that yielded crystals were optimized in 24-well sitting drop plates using 500 µL reservoir wells of varying pH and precipitant concentrations. Crystallization was performed using sitting drop vapor diffusion. Crystallization and crystal soaking were carried out at room temperature. Reflections in all datasets were indexed, integrated, and scaled in HKL2000 (Otwinowski & Minor, 1997). Reciprocal-space refinement was performed using Phenix Refine (Afonine *et al.,* 2012) and the structures were refined in real-space using COOT (Emsley *et al.,* 2010). Waters were added before the final refinement. Final models were evaluated by Molprobity (Chen *et*

al., 2010). All figures were prepared with PyMol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.) and Adobe Illustrator CS4.

General Methods. The PCR amplification of DNA sequences was conducted in a GeneAmp 2700 thermocycler made by Applied Biosystems (Waltham, MA). Techniques for restriction enzyme digestion, ligation, transformation, and other standard molecular biology manipulations were based on methods described elsewhere (Sambrook, 1989). DNA sequencing was performed in the DNA Core Facility in the ICMB. Electrospray ionization mass spectrometry (ESI-MS) was carried out on an LCQ electrospray ion-trap mass spectrometer made by Thermo Scientific (Waltham, MA). Steady-state kinetic assays were conducted on an Agilent 8453 diode-array spectrophotometer at 22 °C. Non-linear regression data analysis was performed using Mathmatica (Wolfram Research, Inc.). Protein concentrations were determined by the Waddell method (Waddell, 1956). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on denaturing gels containing 12% polyacrylamide. Nuclear magnetic resonance (NMR) spectra were recorded on a Varian INOVA-500 or a Varian DirectDrive 600 MHz spectrometer. NMR signals were analyzed using the software program SpinWorks 3.1.6. The sequence alignments and secondary information were visualized using ESPript version 3.0.

2.2.1 Preparation of NahE, PhdJ, and PhdG

Bacterial Strains and Plasmids. The plasmid designated pRE701 (carrying the *nahE* gene) was a gift from Dr. Richard Eaton. The *Mycobacterium vanbaalenii* PYR-1

genomic DNA was a gift from Dr. Carl Cerniglia (National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR). *Escherichia coli* ArcticExpress cells were obtained from Agilent Technologies (Santa Clara, CA).

Construction of the NahE Expression Vector. *NahE* was amplified using the PCR with 5'- TAGTAGTAGCATATGTTGAATAAAG -3' and 5'-GATGATGATCTCGAGTCATTATTATTTACTGTATTTAGCGTG -3', as the forward and reverse primers, respectively. The primers contained the *Xho*I and *Nde*I restriction sites (underlined). The PCR product and the pET24 vector were treated with the appropriate restriction enzymes, ligated, and processed to construct an expression vector for NahE. Three mutations (D265G, S275T, and P299S) were introduced using the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara) following the manufacturer's instructions. These changes corrected mutations that had been introduced during the PCR.

Construction of the PhdJ Expression Vector. The gene, *phdJ* (Mvan_0469) was amplified from the *M. vanbaalenii* PYR-1 genomic DNA using the PCR with 5'-CGAGAGAGCATATGGTGCACGT-3' and 5'-TCCTCAGGATCCGTGGTTCGAGAC-3' as forward and reverse primers, respectively. The primers contained *Nde*I and *Bam*HI restriction sites (underlined). The PCR product and the pET24 vector were treated with the appropriate restriction enzymes, ligated, and processed to construct an expression vector for PhdJ. The PCR introduced two mutations, which were corrected (K44T, and G335Stop) using the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies) following the manufacturer's instructions. The plasmid containing the *phdJ* gene was

transformed into *E. coli* Arctic Express cells (Agilent Technologies) following the manufacture's transformation protocol. The ATG start codon was positioned immediately before the GUG start codon found for *phdJ M. vanbaalenii* PYR-1, so the expressed protein had an additional residue, valine, on the N-terminus.

Construction of PhdG Expression Vector. The *phdG* gene was amplified from *M. vanbaalenii* PYR-1 genomic DNA using the PCR with 5′-CACGGCTCGGATCCACAACACG-3′ and 5′-CCGGCGTGAGCTCCTATGACGAGG-3′ as forward and reverse primers, respectively. The primers contained *Bam*HI and *Sac*I restriction sites (underlined), which were used for insertion into a pET24a vector. QuikChange mutagenesis was performed to correct several mutations including the start codon and Q325E. *E. coli* DH5α cells were used for plasmid replication. A sequence encoding a C-terminal His-tag and TEV cleavage site was also inserted using Gibson Assembly following the manufacturer's protocol (New England Biolabs, Ipswich, MA). When the sequenced DNA showed that no mutations were present in the gene, the plasmid was transformed into Arctic Express cells.

Expression and Purification of NahE. The pET vector construct (encoding *nahE*) was transformed into *E. coli* ArcticExpress cells and used to inoculate LB media (450 mL) containing kanamycin (50 μg/mL), and gentimicin (20 μg/mL). After the cells had been shaken overnight at 37°C, 25 mL of the culture was used to inoculate each of 9 2-L Erlenmeyer flasks containing 500 mL M9 minimal media (Sambrook, 1989) with 0.4% glucose, kanamycin (50 μg/mL), and gentimicin (20 μg/mL). The cells were allowed to grow until the OD_{600} reached ~0.5 (~3 h). In one preparation, the flasks containing the

cells were put on ice for 30 min, then β-D-thiogalactoside (IPTG) was added to make each 1 mM in IPTG. After growing for ~65 h at 15 °C, cells were harvested by centrifugation (20 min at 11300 rpm) and stored at -20 °C. In another preparation, NahE was expressed in LB media and induced overnight with IPTG (1 mM) at room temperature.

In a typical procedure, cells were thawed and suspended in 150 mL of 50 mM sodium phosphate (pH 7.0) containing protease inhibitors [0.5 mM phenylmethylsulfonyl fluoride (PMSF), 5 mM EDTA, 1.5 mg/mL lysozyme]. Cells were lysed by sonication using a W-385 ultrasonicator made by Heat Systems (15 min at 30% duty, 5 s cycles). Ground $(NH_4)_2SO_4$ was added to the lysate to 30% saturation before centrifugation (30 min at 39000 *g*). The supernatant was loaded onto a Phenyl Sepharose 6 (~55 mL, GE) column equilibrated in 50 mM sodium phosphate (pH 7.5) made up to 15% in $(NH_4)_2SO_4$. After the column was washed (>1 column volume, CV), protein was eluted by using a gradient (~200 mL) of 50 mM sodium phosphate (pH 7.0) made 15% in $(NH_4)_2SO_4$ to 50 mM sodium phosphate pH 7.0. The protein did not completely elute from the column as determined by the Bradford assay, so a 100 mL gradient was run from 50 mM sodium phosphate to de-ionized water, which resulted in complete elution. Fractions (150 drops) were collected and analyzed by SDS-PAGE for the appearance of a band around 35000 Da. Fractions were pooled based on the purity (assessed by SDS-PAGE). The protein was exchanged into 50 mM sodium phosphate (pH 7.0) by dialysis overnight.

The dialyzed protein was loaded on a DEAE anion exchange column (~30 mL) equilibrated in 50 mM sodium phosphate (pH 7.0). Protein was eluted using a gradient (~100 mL) to 50 mM sodium phosphate (pH 7.0) made 0.5 M NaCl. Fractions of eluent (~2 mL) were collected and analyzed by SDS-PAGE. Fractions were pooled based on purity (as assessed by SDS-PAGE) and dialyzed into 50 mM sodium phosphate (pH 7.0).

Protein used for crystallography was concentrated and further purified by gel filtration chromatography using an a HiLoad 16/60 Superdex (120 mL, GE) column connected to a fast protein liquid chromatography (FPLC) setup. The protein was eluted isocratically (1 mL/min) in 25 mM HEPES (pH 7.5) made 0.1 M in NaCl. Fractions (1.0 mL) were collected, analyzed by SDS-PAGE, and pooled based on purity. The pooled fractions were concentrated and divided into small aliquots. These stocks were flash-frozen with liquid nitrogen, then stored at -80 °C. Typically the yield for this procedure is ~30 mg per L of culture or 3 mg per g of cells purified. The purity of NahE was determined to be >95% as assessed by SDS-PAGE and electrospray ionization mass spectroscopy (ESI-MS).

Preparation of SeMet-NahE. The L-selenomethionine (SeMet) derivative of NahE was expressed and purified similar to wild type enzyme with the following modifications. Cells were grown in M9 minimal media (Sambrook, 1989) with 0.2% glucose until the OD_{600} reached 0.5-1.0. To each flask were added lysine, phenylalanine, threonine (100 mg each); isoleucine, leucine, and valine (50 mg each); and selenomethionine (60 mg). Additionally, IPTG (1 mM) was added prior to incubation (16 h) at room temperature. The remainder of the purification protocol for the derivative was identical to wild-type

except that dithiothreitol (5 mM) was added to all buffers to prevent oxidation of selenium. The size-exclusion chromatography step was included in the purification.

Preparation of K183A NahE. QuikChange Mutagenesis was performed to generate a plasmid encoding K183A NahE. Once the sequence was confirmed by DNA sequencing, the mutant was expressed and purified similar to wild-type and submitted for ESI-MS.

Expression and Purification of PhdJ, S278N PhdJ, and D282E PhdJ. The *E. coli* Arctic Express cells (containing the PhdJ expression vector) were used to inoculate a starter culture containing 50 mL LB media with kanamycin (50 μg/mL) and gentimicin (20 μg/mL). The cultures were grown at 37°C overnight. Subsequently, 4 2-L Erlenmeyer flasks containing 400 mL ZYM-5052 auto-induction media (Studier, 2005) and kanamycin (100 μg/mL) were inoculated with 25 mL starter culture (each) and grown until the OD_{600} reading was ~0.3 (~3 h). The cultures were then cooled to 12° C and shaken at 250 rpm for ~65 h. Cells were harvested by centrifugation at 11300 rpm for 20 min and stored at -20 °C. The cell pellet mass for the 1.6 L culture was 25.5 g.

The cells were thawed and suspended in 50 mM sodium phosphate (pH 7.0, 80 mL) containing protease inhibitors (1 mM PMSF, 5 mM EDTA, 1.5 mg/mL lysozyme). The cells were lysed by sonication using a W-385 ultrasonicator from Heat Systems (20 min. at 30% duty, output level 6, at 5 s intervals). The lysate was centrifuged (30 min at 12,000 rpm) and the pellet was discarded. 50 mM sodium phosphate buffer (pH 7.0, ~30 mL) was added to the supernatant and the solution was centrifuged again (20 min at 18,000 rpm). This supernatant was loaded onto a DEAE column (17 mL) equilibrated

with 50 mM sodium phosphate buffer (pH 7.0). The column was washed (>1 CV) with the same buffer. Protein was then eluted with a linear gradient (0-0.5 M NaCl in the sodium phosphate buffer, 85 mL). Fractions (1.4 mL) were pooled based on the presence of an SDS-PAGE band at ~35 kD and activity using substrate, *ortho*-carboxybenzylidenepyruvate (CBP).

$(NH_4)_2SO_4$ (30%) was added to the pooled fractions and stirred for 20 min on ice for ~20 min prior to centrifugation (20 min at 18,000 rpm). After storage of the supernatant overnight at 4 °C, $(NH_4)_2SO_4$ (40%) was added. The colution was stirred for ~20 min on ice before centrifugation (20 min at 18,000 rpm). The pellet was dissolved and dialyzed into 50 mM sodium phosphate (pH 7.0). For a typical preparation, the yield was 4.9 mg per L culture or 0.3 mg per g cells. An additional hydrophobic interaction column using Phenyl Sepharose column was employed when the purity was <95% as assessed by SDS-PAGE. The purity and identity of PhdJ was determined by SDS-PAGE and ESI-MS, respectively. The homogeneous protein was divided into small aliquots (50-200 μL), flash-frozen with liquid nitrogen, and then stored at -80°C.

Expression and Purification of PhdG. All culture media contained kanamycin (50 μg/mL) and gentimicin (20 μg/mL). For expression of PhdG, LB media (7 mL) was inoculated with Arctic Express cells containing pET24a (encoding *phdG*). After the cells were grown 37 °C overnight with shaking, LB media (50 mL) was added, and the culture was incubated for 4 h. Culture (3 mL) was used to inoculate each of 12 2-L flasks containing 400 mL of ZYM-5052 media (Studier, 2005). The cultures were incubated for 2.5 h with shaking at 37 °C until the OD_{600} reached 0.6. The culture was then incubated

with shaking at 12 °C for ~65 h prior to centrifugation (20 min at 8000 rpm). The cell pellet (95 grams) was stored at -80 C°.

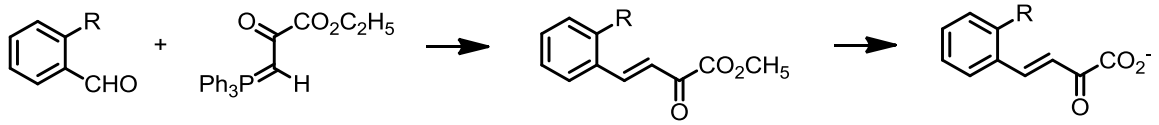
Cells were thawed and suspended (by stirring) in 50 mM sodium phosphate (pH 7.0, ~200 mL) containing EDTA (5 mM), PMSF (1 mM), and lysozyme (1 mg/mL). The cells were lysed by sonication using a W-385 ultrasonicator from Heat Systems (20 min. at 30% duty, output level 6, at 5 s intervals) prior to centrifugation of the lysate (20 min at 20,000 rpm). The pellet was discarded and the supernatant was then centrifuged again (30 min at 60,000 rpm), and loaded on a HisPur column (8 mL) equilibrated with 50 mM sodium phosphate (pH 7.0) made 300 mM in NaCl. The column was washed with the equilibration buffer (5 CV). To elute PhdG, a linear gradient (50 mL) was used from the equilibration buffer to 50 mM sodium phosphate (pH 7.0) containing 300 mM NaCl and 200 mM imidazole. Fractions (~1.6 mL) were collected and analyzed by SDS-PAGE. Fractions that appeared to be most homogeneous were treated with TEV protease (600 µL, 1.8 mg/mL, gift from Dr. Tyler Stack). The digested PhdG solution was then dialyzed into 25 mM HEPES (pH 7.5) containing 0.1 M NaCl (4 L, 6000 Da MW cutoff) and concentrated using an Amicon to 5.8 mg/mL.

Buffer-exchanged PhdG (5 mL) was injected onto a Hiload 16/60 Superdex (GE) size-exclusion column and eluted isocratically (1 mL/min) in 25 mM HEPES (pH 7.5) containing 0.1 M NaCl. Fractions (1.0 mL) were collected and analyzed by SDS-PAGE. Fractions were combined based on homogeneity (assessed by SDS-PAGE) and concentrated. Concentrated PhdG was divided into small aliquots (100 and 200 µL) and

flash frozen using liquid nitrogen prior to storage at -80 °C. The typical yield for PhdG was 24 mg per L culture or 1.2 mg per g of cell pellet.

2.2.2 Synthesis

trans-o-Hydroxybenzylidenepyruvate (HBP). Salicaldehyde (1 g, 13 mmol) was added to ethyl triphenylphosphoranylidenepyruvate (5 g, 13 mmol, Ernest *et al.*, 1978, Johnson *et al.*, 1995) in anhydrous DMF with a catalytic amount of benzoic acid (50 mg) and was heated at 80°C for ~15 h (**Scheme 2-1**). Water was added (25 mL) prior to extraction with hexanes-ethyl acetate (2:1). Na_2SO_4 was added to the organic phase and the solution was run through a plug of silica. Silica chromatography using hexanes-ethyl acetate (4:1) was performed. Fractions were collected and evaporated to dryness. The residue was redissolved in H_2O and made alkaline by adding NaOH (1 M) dropwise so as to not exceed pH 10. The solution was extracted with ethyl acetate to remove ethanol and remaining ethyl ester. The aqueous phase was then acidified to pH 2 and the product was isolated by extraction with ethyl acetate. The ethyl acetate layers were dried using anhydrous Na_2SO_4 and evaporated to dryness. The yield was 1 g (64%). **HBP**: ^{1}H NMR (H_2O, 600 MHz) δ 6.773 (1H, d, J = 16.44 Hz), 6.786 (1H, d, J = 7.56 Hz), 6.805 (1H, t, J = 7.86 Hz), 7.196 (1H, td, J = 7.77 Hz, 1.65 Hz), 7.469 (1H, dd, J = 7.86 Hz, 1.62 Hz), 7.789 (1H, d, J = 16.50 Hz)].



Scheme 2-1 Synthesis of HBP (R=OH) and CBP ($R=CO_2^-$), starting with salicylaldehyde (R=OH) or CBA ($R=CO_2^-$).

ortho-Carboxybenzylidenepyruvate (CBP). *ortho*-Carboxybenzaldehyde (CBA) methyl ester was synthesized by mixing *o*-carboxybenzaldehyde (1.5 g, 10 mM) and dimethyl sulfate (1.9 g, 15 mM) with ~2.5 equivalents K_2CO_3. The reaction was stirred at a gentle reflux in acetone (50 mL). The majority of the acetone is removed under reduced pressure. The residue is diluted with 100 mL ethanol and filtered to remove salts. The ethyl acetate is washed with water, dried ($NaSO_4$), and the solvent is removed to give the methyl ester. Water was added and the methyl ester was extracted with hexanes-ethyl acetate (4:1). The organic phase was dried over anhydrous Na_2SO_4 and evaporated to dryness. A mixture of 2-carboxybenzaldehyde methyl ester (1 g, 13 mmol), ethyl triphenylphosphoranylidenepyruvate (5 g, 13 mmol, Ernest *et al.*, 1978, Johnson *et al.*, 1995) and a catalytic amount of benzoic acid (50 mg) was heated in anhydrous DMF under argon at 80°C (~15 h) (**Scheme 2-1**). Water (25 mL) was added and the reaction mixture was extracted with hexanes-ethyl acetate (2:1). The organic layers were combined, dried over anhydrous Na_2SO_4, and run through a small silica plug. The compound was purified using flash chromatography (hexanes-ethyl acetate, 4:1). After removing the solvent under reduced pressure, CBP methyl ester was redissolved in H_2O (~5 mL) and hydrolyzed by NaOH (1 M), dropwise so as to not exceed pH 10. The

solution was extracted with ethyl acetate. The aqueous phase was then rapidly acidified (pH 2) using 8.5% phosphoric acid. CBP was extracted using ethyl acetate, dried over anhydrous Na_2SO_4, and evaporated to dryness. The overall yield was ~1 g. **CBP**: ^{1}H NMR (H_2O, 600 MHz) δ 6.719 (1H, d, J = 16.28 Hz), 7.268-7.351 (3H, m), 7.620 (1H, d, 16.34 Hz), 7.864 (1H, d, J = 16.34).

2-*hydroxychromene*-2-*carboxylate (HCCA).* HCCA was prepared by dissolving HBP in 100 mM dibasic sodium phosphate (200μL) and water (400 μL). NaOH (25 M) was added in 3 μL increments until the pH was above 5, then NaOH (1 M) was added in 3 μL increments until the pH was 7. The solution was incubated overnight to form a mixture HBP and HCCA in equilibrium. The solution was loaded onto a G-15 sepharose column (~7 mL). Fractions were collected on ice and fractions with an A_{255}/A_{296} greater than 2.5 were pooled and determined to be adequately free of the *trans* isomer by ^{1}H NMR spectroscopy. The fractions were flash frozen and stored at -20°C. **HCCA:** ^{1}H NMR (H_2O, 600 MHz) δ 5.764 (1H, d, J = 9.74 Hz), 6.658 (1H, d, J = 9.76 Hz), 6.821 (1H, d, J = 8.07 Hz), 6.873 (1H, t, J = 7.309 Hz), 7.0943 (1H, d, J = 7.46 Hz), 7.128 (1H, t, J = 7.71 Hz).

Synthesis of trans-Benzylidenepyruvate (BP). BP was synthesized using a published procedure (Reimer, 1931). **BP:** ^{1}H NMR (H_2O, 600 MHz) δ 6.717 (1H, d, J = 16.51 Hz), 7.285-7.351 (3H, m), 7.514 (1H, d, 16.51 Hz), 7.536-7.508 (2H, m).

2.2.3 Steady-State Kinetics

Determination of Extinction Coefficients for HBP, salicylaldehyde, HCCA, CBP, CBA, BP, trans-Cinnamaldehyde, and 2-oxo-6-phenyl-3,5-Hexadienoate. To determine the ε_{296} for HBP, 5.2 mg was dissolved in ethanol (10 mL). An aliquot (25 μL) of this stock was then added to 50 mM sodium phosphate (970 μL, pH 7.0). The ε_{296} for salicylaldehyde was determined by dissolving 6.0 mg in 50 mM sodium phosphate (3.00 mL, pH 7.0). An aliquot (5 μL) of this solution was diluted in 50 mM sodium phosphate (995 μL, pH 7.0). The ε_{300} for CBP was determined by measuring the absorbance of CBP (81 μM) in 50 mM sodium phosphate (pH 7.0). To measure ε_{300} for CBA, 3.6 mg was dissolved in 1.00 mL sodium phosphate (pH 7.0). An aliquot (5 μL) was diluted into 50 mM sodium phosphate (995 μL, pH 7.0). To measure ε_{300} for BP, 9.7 mg was dissolved in ethanol (1.0 mL). The stock was diluted 10-fold with ethanol, and an aliquot (5 μL) was added to 50 mM sodium phosphate (1.0 mL, pH 7.0). The ε_{340} for *trans*-cinnamaldehyde was determined by adding *trans*-cinnamaldehyde (5 μL, 2.05 mg/mL) to 50 mM sodium phosphate (980 μL, pH 7.0). To determine the ε_{340} 2-oxo-6-phenyl-3,5-hexadienoate, which was not isolated, pyruvate (5 μL, 325 mg/mL) and PhdG (20 μL, 8.4 mg/mL) were then added to the solution with *trans*-cinnamaldehyde to force its complete conversion to 2-oxo-6-phenyl-3,5-hexadienoate. The reaction was assumed to be complete when the absorbance did not change significantly after a second addition of pyruvate (5 μL, 325 mg/mL). All measurements were performed at room temperature.

Non-enzymatic Isomerization of HBP and HCCA. Isomerization was measured in 50 mM sodium phosphate buffer (pH 7.0) by UV absorbance every min for 60 min. The extinction coefficients for HCCA at wavelengths of 255, 296, and 340 nm were 6800 $M^{-1}cm^{-1}$, 2700 $M^{-1}cm^{-1}$, and 700 $M^{-1}cm^{-1}$ respectively. The extinction coefficients for HBP at wavelengths of 255, 296, and 340 nm were determined to be 4000 $M^{-1}cm^{-1}$, 11000 $M^{-1}cm^{-1}$, and 7000 $M^{-1}cm^{-1}$ respectively. Non-enzymatic rates were calculated by dividing absorbance changes by the absorbance differences of the species. All measurements were carried out at room temperature.

Steady-State Kinetic Analysis of NahE with HBP. HBP (15.3 mg) was dissolved in ethanol (1.4 mL) and diluted to make separate stocks of 2.19, 1.56, 1.09, 0.728, 0.437, 0.273, 0.168, 0.0397, 0.0546, 0.0728 mg/mL. NahE (6.5 mg/mL) in 50 mM sodium phosphate buffer (pH 7.0) was diluted to 23 μg/mL (0.63 μM) for use in the enzyme assays. Substrate stock (10 μL) was added to 50 mM sodium phosphate (985 μL, pH 7.0) in a 1 mL cuvette. The reaction was initiated by the addition of NahE (5 μL) to a final concentration of 3.2 nM (0.12 μg/mL). Final substrate concentrations in the reactions ranged from 1.51-114 μM. Reaction solutions were mixed by inversion. Reactions were monitored at 296 nm for 150 s using a 10 s cycle for concentrations 114-3.79 μM, 75 s with a 5 s cycle for 2.84 and 1.94 μM concentrations, and 45 s with a 3 s cycle for the 1.51 μM concentration. The difference in ε_{296} for HBP (11,000 $M^{-1}cm^{-1}$) and salicylaldehyde (1300 $M^{-1}cm^{-1}$) was 9700 $M^{-1}cm^{-1}$. This value was used to convert initial rates from AU/s to M/s, which were then converted to k_{obs} for fitting to the following equation that accounts for substrate (HBP) inhibition:

$$k_{\mathrm{obs}} = \frac{k_{\mathrm{cat}}[\mathrm{HBP}]}{K_{\mathrm{M}} + [\mathrm{HBP}]\left(1 + \frac{[\mathrm{HBP}]}{K_{\mathrm{I}}^{\mathrm{HBP}}}\right)} \quad (\mathbf{1})$$

All measurements were carried out at room temperature.

NahE Activity with HBP in the Presence of Salicylaldehyde. An aliquot of HBP (10 µL, 15, 9.0, 5.4, 3.0, 1.8, 1.1 mM) was added to initiate 1.0 mL reactions containing 50 mM sodium phosphate (pH 7.0), NahE (18 nM), and various concentrations of salicylaldehyde (10 µL 10.1, 5.07, 2.53, 1.26, 0.633, and 0.317 mM dissolved in ethanol were added to give final concentrations of 101, 50.7, 25.3, 12.6, 6.33, and 3.17 µM HBP). The final concentration range of HBP was 11-150 µM. Initial rates were measured by A_{296} decrease over 25 seconds with spectra recorded every 2.5 seconds. All measurements were carried out at room temperature.

Steady-State Kinetics of NahE with BP and CBP. Stock solutions of BP (1.3-11 mM) were prepared in ethanol. Stock solutions of CBP (0.525-52.5 mM) were prepared in 50 mM sodium phosphate (pH 7.0). Reactions were initiated by the addition of NahE (final concentrations of 1.5–3.4 µM). Reactions were monitored by the decrease in absorbance decrease at 300 nm. Reactions were monitored after initiation for 30-120 s with a 2.5-6 s cycle time depending on the reaction rate. The ε_{300} for substrates BP and CBP were 21000 $M^{-1}cm^{-1}$ and 10200 $M^{-1}cm^{-1}$, respectively. The ε_{300} for the products benzaldehyde and CBA were 540 $M^{-1}cm^{-1}$ and 1450 $M^{-1}cm^{-1}$, respectively. The difference in ε_{300} for substrates and products were used to calculate k_{obs} which were

plotted vs. substrate concentration and fit to the Michaelis-Menten equation. All measurements were carried out at room temperature.

Relative Activity of K183A NahE. K183A NahE was made 110 nM to initiate a 1.00 mL reaction with HBP (84 μM) in 50 mM sodium phosphate (pH 7.0). The same reaction was performed using wild-type NahE (110 nM). The intitial decrease of A_{296} was measured. All measurements were carried out at room temperature.

NahE Activity with Salicylaldehyde and Pyruvate. Stock solutions of pyruvate (20 μL) and salicylaldehyde dissolved in ethanol (5 μL, diluted from a 2 M stock) was added to 50 mM sodium phosphate (pH 7.5, 965 μL) to reach final concentration ranging from 25 μM to 1.0 mM and 0.4 to 40 mM, respectively. NahE (10 μL of a 0.13 mg/mL solution) was added and mixed in the solution to a final concentration of 35 nM to initiate reactions. Spectra were recorded every 3 s over 30 min. Initial rates were calculated using increase of A_{296} and were fit to the following equation for a ping-pong mechanism:

$$k_{\mathrm{obs}} = \frac{k_{\mathrm{cat}}[\mathrm{pyr}][\mathrm{ald}]}{K_{\mathrm{pyr}}[\mathrm{ald}] + [\mathrm{pyr}][\mathrm{ald}] + K_{\mathrm{ald}}[\mathrm{pyr}]} \quad (2)$$

where ald is salicylaldehyde, and pyr is pyruvate. All measurements were carried out at room temperature.

PhdJ Activity with CBP. CBP (2.0 mg, disodium salt) was added to 50 mM sodium phosphate buffer (pH 7.0, 500 μL) to make a 16 mM solution (pH ~8 by pH paper). The stock solution was diluted to make various concentrations (8.1 mM, 4.1 mM,

2.4 mM, 1.6 mM, 1.1 mM, and 0.81 mM) in the same sodium phosphate buffer. Subsequently, aliquots (10 μL) of the diluted substrate stock solutions were added to 50 mM sodium phosphate buffer (985 μL, pH 7.0). PhdJ (5 μL of a 0.51 mg/mL solution) was added and mixed, to a final concentration of 71 nM, with CBP (81, 41, 24, 16, 11, and 8.1 μM) to initiate reactions. Initial rates were measured by following the decrease of absorbance at 300 nm. The initial rates were converted to molar units using the difference between extinction coefficients for CBP (10500 $M^{-1}cm^{-1}$) and CBA (1450 M^{-1} cm^{-1}). The activities for the S278N and D282E mutants of PhdJ were determined by adding CBP (7 μL, 0.489-2.75 mg/mL) to 50 mM sodium phosphate (pH 7.0, 986 μL). Reactions were initiated by adding S278N and D282E mutants of PhdJ (10 μL made 33.7 μM and 100 μL made 15.5 μM, respectively, in sodium phosphate, pH 7.0). The k_{cat}/K_M values for these reactions were determined by a linear fit of k_{obs} plotted vs. CBP concentration with the y-intercept set to 0 (k_{obs} is 0 without substrate). All measurements were carried out at room temperature.

PhdJ, S278N PhdJ, and D282E PhdJ Activity with HBP. The substrate (2.7 mg) was dissolved in ethanol (1.0 mL) and further diluted (1.8, 1.4, 0.90, 0.54, and 0.36 mg/mL) in ethanol. Aliquots (6 μL) of the diluted substrate stocks were added to 50 mM sodium phosphate buffer (942 μL, pH 7.0) so that the final concentrations of HBP were 85, 56, 42, 28, 17, and 11 μM. The reaction was initiated by adding 50 μL aliquots of a PhdJ solution (0.51 mg/mL) to make a final concentration of 0.71 μM. Initial rates were measured by following the decrease of absorbance at 296 nm. Spectra were recorded every 5 s for 60 s. Initial rates were calculated using the differences in the ε_{296} for HBP

(10700 $M^{-1}cm^{-1}$) and salicylaldehyde (1260 $M^{-1}cm^{-1}$). Activities for the S278N and D282E PhdJ mutant were determined by adding HBP (7 µL, 0.293-2.20 mg/mL in ethanol) to 50 mM sodium phosphate (pH 7.0, 986 µL). Reactions were initiated by adding S278N and D282E PhdJ mutant (7 µL made 33.7 µM and 20 µL made 15.5 µM, respectively, in sodium phosphate, pH 7.0) was added. The k_{cat}/K_M values for these reactions were determined by a linear fit of k_{obs} plotted with HBP concentration setting the y-intercept to 0. All measurements were carried out at room temperature.

PhdJ, S278N PhdJ, and D282E PhdJ Activity with BP. BP (5.4 mg) was dissolved in ethanol (1 mL) and further diluted to 1.1, 0.72, 0.54, 0.36, 0.27, 0.18, and 0.12 mg/mL in ethanol. Aliquots (7 µL) of the diluted stock solution were added to 50 mM sodium phosphate buffer (pH 7.0, 943 µL) so that final concentrations of BP were 48, 32, 24, 16, 12, 7.9, and 5.3 µM. The reactions were initiated by adding PhdJ (10 µL, 0.51 mg/mL) to a final concentration of 1.4 µM. Spectra were recorded every 20 s for 10 min. Initial rates were determined by following the decrease of absorbance at 300 nm. The ε_{300} for BP and benzaldehyde was determined to be 21,000 ± 1,000 and 540 ± 20 M^{-1} cm^{-1}, respectively. The activity for S278N and D282E PhdJ mutants were measured by adding BP (7 µL, 0.128-1.01 mg/mL in ethanol) to 50 mM sodium phosphate (pH 7.0, 973 µL). The reactions were intitiated by the addition of S278N and D282E PhdJ mutants (20 µL, 33.7 µM and 15.5 µM, respectively). All measurements were carried out at room temperature.

PhdG Activity with trans-Cinnamaldehyde and Pyruvate. trans-Cinnamaldehyde (17.7 mg) was dissolved in ethanol (981 µL) to make a 17.7 mg/mL stock. This stock solution was diluted to make solutions ranging from 0.443 mg/mL to 17.7 mg/mL.

Sodium pyruvate (17.5 mg) was dissolved in 50 mM sodium phosphate (pH 7.0, 318 μL), and was diluted to make 7.0, 1.75, and 0.70 mg/mL of pyruvate stocks. Using these solutions, *trans*-cinnamaldeyde (5 μL) and pyruvate (5 μL) were added to 50 mM sodium phosphate buffer (pH 7.0, 985 μL). PhdG (5 μL, 0.17 mg/mL) was added to initiate reactions. Initial rates were measured by following the decrease of absorbance at 340 nm measured every 3 s for 60 s. The ε_{340} for cinnamaldehyde and the product were determined to be 570 $M^{-1}s^{-1}$ and 14400 $M^{-1}s^{-1}$. Rates were plotted vs. pyruvate and cinnamaldehyde concentration and fit to **eq. 2**. All measurements were carried out at room temperature.

Steady State Kinetics for PhdG using BP. BP (12.1 mg) was dissolved in ethanol (1.0 mL) and diluted into stock solutions ranging from 0.128-1.01 mg/mL. An aliquot (7 μL) was added to 50 mM sodium phosphate buffer pH 7.0 (986 μL). To initiate reactions, PhdG (7 μL, 1.7 mg/mL) was added. Spectra were recorded every 3 s for 60 s. Initial rates were determined by following the decrease of absorbance at 300 nm. All measurements were carried out at room temperature.

2.2.4 ^{1}H NMR Spectrometric Time Course for Reactions Catalyzed by NahE, NahD, PhdJ, and PhdG

^{1}H NMR Time Course of Reactions Catalyzed by NahE and NahD. HBP or HCCA (4.0 mg, 30 mM) dissolved in DMSO-d_6 (30μL) was added to 50 mM sodium phosphate (600 μL, pH 7.5) to monitor the production of aldehyde and pyruvate. Reactions with

benzaldehyde and salicyaldehyde were prepared in a like manner, but pyruvate (1 eq) dissolved in the sodium phosphate buffer was added first to the NMR tube. Reactions were initiated by adding NahE (100 μL, 4.1 mg/mL) and/or NahD (100 μL, ~1 mg/mL). Spectra were recorded every 3 min for 33 min. The NMR spectra were recorded on a Varian INOVA-500 spectrometer (Palo Alto, CA), using selective presaturation of the water signal with a 2 s presaturation interval. The lock signal is DMSO-d_6. Chemical shifts are standardized to the DMSO-d_6 signal at 2.49 ppm. Reactions prepared for the measurement of NahE activity were: HBP only, HCCA only, NahD and HBP, benzaldehyde and pyruvate, and salicylaldehyde and pyruvate. Control samples were prepared in a like manner, but buffer (100 μL) was added instead of enzyme. Approximate amounts of product were determined by integration.

^{1}H NMR Identification of the Products in the PhdJ-catalyzed Reaction. To NMR tubes containing 100 mM sodium phosphate buffer (600 μL, pH 7.5) was added a quantity of CBP or CBA (4.5 mg in 30 μL DMSO-d_6). For the reaction containing CBA, pyruvate (1 eq) was dissolved in the same buffer that was initially added to the NMR tube. The reaction was initiated by adding PhdJ (100 μL of a 0.51 mg/mL solution) to a final concentration of 1.9 μM. The final pH of the reaction mixtures were 7.5. DMSO-d_6 was used as the lock signal and for the standardization of the chemical shifts (at 2.49 ppm). A non-enzymatic control was prepared identically except that PhdJ was not included in the reaction. NahE spectra were recorded every 3 min over the course of 33 min.

^{1}H NMR Identification of the Products in the PhdG-catalyzed Reactions. To an NMR tube containing 100 mM sodium phosphate buffer (600 µL, pH 7.5) was added a quantity of BP or cinnamaldehyde (4.5 mg in 30 µL DMSO-d_6). For the reaction containing cinnamaldehyde, pyruvate (1 eq) was initially dissolved in the NMR tube using the same buffer. Reactions were initated by adding PhdG (25 µL, 8.3 mg/mL).

2.2.5 UV/Vis Absorbance and Mass Spectrometry of Modified Hydratase-Aldolases

UV/Vis spectra of NahE and PhdJ with benzaldehyde, salicylaldehyde and CBA. Each aldehyde (5 µL, 3.46 mM) was added separately to solutions containing NahE or PhdJ (95 µL, 91 µM) to make final concentrations of 173 µM aldehyde and 86.5 µM NahE or PhdJ (2:1 ratio). The spectrophotometer was blanked immediately after mixing. UV/Vis spectra were monitored over the course of several hours.

NahE and K183A NahE Modification with HBP, salicylaldehyde, pyruvate, 2-oxo-3-Pentynoate (2-OP), and 4-phenyl-2-oxo-3-Butynoate (POB). All modification experiments were performed at room temperature. $NaCNBH_3$ was added to all solutions except the sample containing POB. Stock solutions of $NaCNBH_3$ were made to 0.75 M by dissolving $NaCNBH_3$ in 100 mM dibasic Na_2HPO_4. HBP, salicylaldehyde, pyruvate, 2-OP, and POB (final concentrations of 1.8, 0.29, 8.0, 0.21, and 0.088 mM, respectively) were added to 50 mM sodium phosphate (pH 7.0) containing enzyme (8.9-140 µM) and $NaCNBH_3$ (15 mM). The compounds and additional $NaCNBH_3$ (15-250 mM, final) were added incrementally over the course of several hours until enzyme activity using HBP

had decreased more than 50% relative to the control with $NaCNBH_3$ treatment only (with the exception of pyruvate, where the activity decreased ~30%).

Covalent Modification of PhdJ by CBP, CBA, or Pyruvate in the Presence of $NaCNBH_3$. Three reaction mixtures containing CBP (260 μM), CBA (550 μM), or pyruvate (20 mM) were made up in 50 mM sodium phosphate (pH 7.0) containing $NaCNBH_3$ (~15 mg/mL in 100 mM Na_2HPO_4) and PhdJ (0.5 mg/mL). CBP, CBA, and pyruvate were added to the samples (1-15 μL increments) until enzyme activity had decreased more than 50% (2-4 h). Inactivation using pyruvate was slow compared to the samples with CBP and CBA, and activity only dropped ~30% over 4 h.

Treatment of PhdG with CBP, 2-OP, or POB in the Presence of $NaCNBH_3$. Three reaction mixtures containing CBP (460 μM), 2-OP (460 μM), POB (460 μM), and pyruvate (460 μM) were made in 50 mM sodium phosphate (pH 7.0, 120-160 μL) containing $NaCNBH_3$ (~15 mg/mL in 100 mM Na_2HPO_4) and PhdJ (6 mg/mL).

Preparation of Inactivated NahE and PhdJ for ESI-MS. NahE and PhdJ were exchanged into 10 mM ammonium carbonate (pH 8.2) using PD-10 columns. Samples were concentrated with Amicon Ultra 0.5 mL Centrifugal Filters (3 kD cutoff). Membranes were rinsed first by filling and decanting the cells five times with filtered water, then spinning the cell filled with water at 10,000 rpm for 5 min. The rinsing procedure was performed five times before ESI-MS. The samples were made ~0.1 mg/mL and submitted for ESI-MS.

2.2.6 Dynamic Light Scattering Analysis of NahE, PhdJ, and PhdG

Dynamic Light Scattering. A DAWN HELEOS-II Multi-Angle Static Light Scattering (MALS) instrument (664 nm beam) connected to an Shimadzu LC-20AD HPLC System equipped with a size-exclusion column Model WTC-030S5 (7.8 × 300 mm, pore size 300 Å, Wyatt Technologies) was used to collect data. Molar mass moments in g/mole were determined with 100 μL samples of NahE, PhdJ, and PhdG (25 μM each). The HPLC solvent was 50 mM sodium phosphate buffer (pH 7.0), and the flow rate was 0.4 mL/min.

2.2.7 Crystallization, Soaking, and Structure Determination of NahE

Crystallization and Crystal Soaking. The diffraction-quality crystals of SeMet NahE (**NahE-SeM**) were obtained in 0.1 M MES (pH 6.5-8.0) and 21-26% PEG 2000 monomethyl ether (MME). The crystals were set up in sitting drop and incubated at ~20°C before harvesting for data collection. Crystals were cryoprotected by 25% glycerol prior to vitrification in liquid nitrogen.

To obtain the NahE structure with (*R*)-4-hydroxy-4-(2-hydroxyphenyl)-2-iminobutanoate bound (**NahE-HN**), an apo crystal of NahE was soaked in the mother liquor containing HBP (5 mM) for ~30 min. For obtaining the structure of NahE in complex with salicylaldehyde and pyruvate (**NahE-SP**), an apo crystal was soaked for 19 min in HBP. For the solution of NahE in complex with BP (**NahE-BP**), an apo crystal of NahE was soaked with 10 mM BP in mother liquor overnight (~20 h). To obtain the

crystal structure of NahE modified by salicylaldehyde at Lys-183 (**NahE-Sal**), a crystal was soaked in 10 mM salicylaldehyde in mother liquor for 1 h. All crystals were cryoprotected by 20-25% glycerol in mother liquor prior to vitrification in liquid nitrogen.

Data Collection and Structure Determination. X-ray diffraction data for the NahE-SeM were collected at Advanced Light Source beamline 8.2.1 (ALS, Berkeley, CA) where diffractions in multiple wavelengths were captured. Datasets used for the solution of NahE-HN and NahE-SP were collected using ALS beamlines 5.0.3 and 5.0.1. Data for NahE-Sal and NahE-BP structures were collected at ALS beamline 5.0.2. Statistics for data collection is summarized in **Table 2-9** and **Table 2-10**.

Data were indexed, integrated, and scaled using HKL-2000 (Otwinowski & Minor, 1997). Phases for Nah-SeM were calculated using Autosol from the Phenix suite of programs (Adams et al., 2004) using multiwavelength anomalous dispersion (MAD). The final model is refined to 2.0 Å resolution and used as the search model for the other complex structures in molecular replacement using Phaser (McCoy, 2007) and Autobuild (Terwilliger *et al.*, 2008). Structures were refined using Phenix Refine (Afonine et al., 2012). Manual model building was done using COOT (Emsley *et al.*, 2010). Model structures were evaluated during and after refinement using Molprobity (Chen *et al.*, 2010).

2.2.8 Crystallization, Soaking, and Structure Determination of PhdJ

Crystallization and Crystal Soaking. The two crystallization conditions that gave diffraction-quality crystals of PhdJ were 0.1 M MES (pH 7.0-7.5) with 22-25% PEG 3350 and 0.1 M MES (pH 6.5-7.5) with 20-21% PEG 6000. The former condition was used for structures of native PhdJ (**PhdJ**) and PhdJ in complex with substrate CBP (in two active sites) and products CBA and pyruvate (in the other two active sites, **PhdJ-CBP**), and the latter condition was used in structure of PhdJ in complex with the (*R*)-2-(3-carboxy-1-hydroxy-3-oxopropyl)benzoate intermediate (**PhdJ-CN**) and the PhdJ structure bound to product *o*-carboxybenzaldehyde (**PhdJ-CBA**). Crystals were grown using sitting drop vapor-diffusion method at ~20°C.

For soaking of crystals for the determination of PhdJ-CBP, CBP (1.2 mg) was dissolved in 45 μL dibasic sodium phosphate buffer (100 mM at pH 8-9) and diluted to 10 mM in mother liquor. An apo PhdJ crystal was soaked for 4 min in this solution. The crystal used to obtain the structure of **PhdJ-CN** was soaked for 30 s in the same CBP stock (10 mM) in 0.1 M MES (pH 7.5) made 21% in PEG 6000. The crystal used to obtain **PhdJ-CBA** was soaked with 10 mM CBA in mother liquor for 60 min. The crystals were cryoprotected by 20-25% glycerol in mother liquor.

Data Collection and Structure Determination. X-ray diffraction data were collected at ALS beamline 5.0.3. Data were indexed, integrated, and scaled using HKL-2000 (Otwinowski & Minor, 1997). The native structure of PhdJ was determined by molecular replacement using NahE-SeM as a search model. Complex crystal structures

were solved using the native PhdJ structure as a model. Molecular replacement was performed using Phaser (McCoy, 2007) and Autobuild (Terwilliger *et al.*, 2008). Structures were refined using Phenix Refine (Afonine et al., 2012). Manual model building was done using COOT (Emsley *et al.*, 2010). Model structures were evaluated during and after refinement using Molprobity (Chen *et al.*, 2010). The statisitics for data collection and refinement is summarized in **Table 2-11**.

2.2.9 Crystallization, Soaking, and Structure Determination of PhdG

Crystallization and Crystal Soaking. The diffraction-quality crystals of PhdG were obtained in 0.1 M MES (pH 7.0) and 19% PEG 6000. Crystals formed using sitting drop at 20°C. The structure of PhdG bound with pyruvate carbinolamine (**PhdG-Pyr**) was soaked in 50 mM sodium phosphate buffer (pH 7.0) made 50 mM sodium pyruvate prior to cryo-protection by 20-25% glycerol in mother liquor for 30 min.

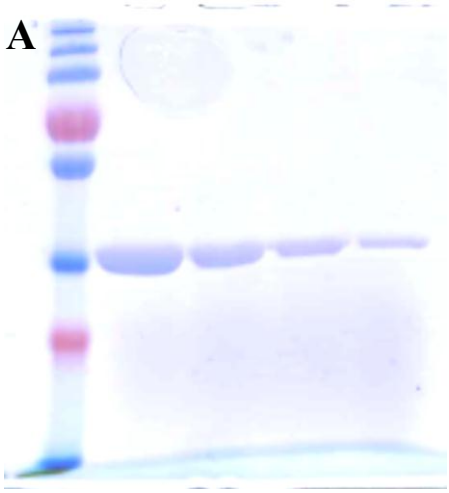
Data Collection and Structure Determination. X-ray data for PhdG were collected at Advanced Light Source beamline 5.0.3 (ALS, Berkeley, CA). Data were indexed, integrated, and scaled using HKL-2000 (Otwinowski & Minor, 1997). The native PhdG structure was determined by molecular replacement using an apo NahE as a search model, and the refined PhdG structure was used as a search model for for PhdG-Pyr.The apo structure of PhdG determined was used as a search model in the solution PhdG-Pyr. Molecular replacement was performed using Phaser (McCoy, 2007) and Autobuild (Terwilliger *et al.*, 2008). Structures were refined using Phenix Refine (Afonine et al., 2012). Manual model building was done using COOT (Emsley *et al.*, 2010). Model

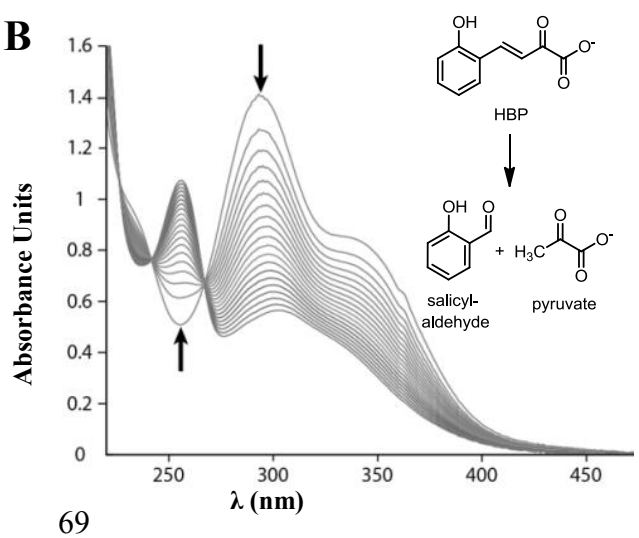
structures were evaluated during and after refinement using Molprobity (Chen *et al.*, 2010). The statisitics for data collection and refinement is summarized in **Table 2-12**.

2.3 RESULTS

2.3.1 Purification and Enzyme Activity of NahE, PhdJ, PhdG

NahE, SeMet NahE, K183A NahE, PhdJ, S278N PhdJ, D282E PhdJ, and PhdG were purified to homogeneity, as determined by SDS-PAGE (NahE shown in **Figure 2-1A**) and ESI-MS (**Section 2.3.5**). Peaks observed by ESI-MS were consistent with expected monomeric mass values. NahE, PhdJ, and PhdG converted HBP to salicylaldehyde and pyruvate, as indicated in the UV-absorbance spectra, where the decrease at 296 nm (HBP peak) accompanies an increase at 255 nm (salicylaldehyde peak). All reactions in this section were performed in 50 mM sodium phosphate buffer (pH 7.0) unless otherwise indicated.





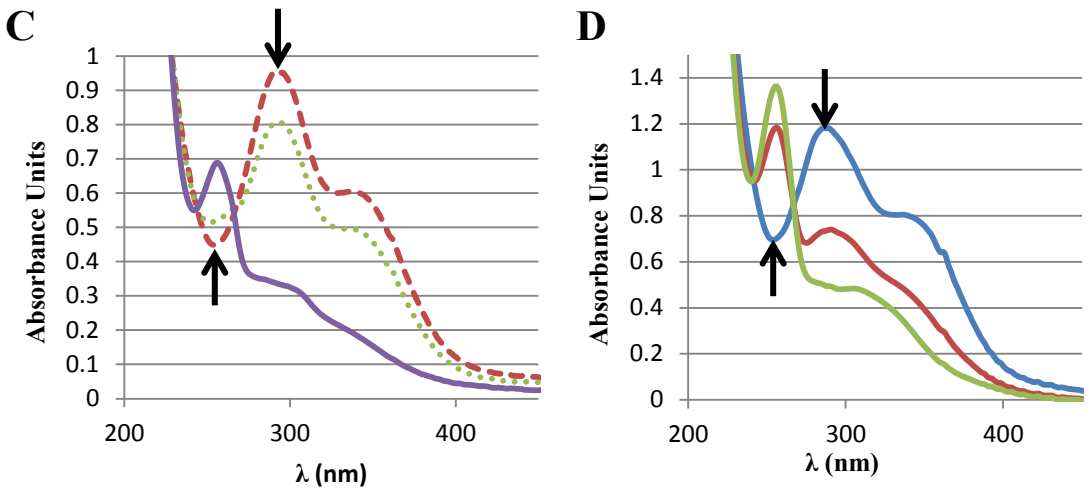


Figure 2-1 Purified NahE and absorbance spectra of HBP in the presence of NahE, PhdJ, and PhdG.

A) SDS PAGE of purified NahE. Lane 1: PageRuler Plus Protein Marker, MW from top (kDa): 250, 130, 100, 70, 55, 35, 15. Lane 2: 8 µg NahE. Lane 3: 4 µg NahE. Lane 4: 2 µg NahE. Lane 5: 1 µg NahE. **B)** HBP (120 µM) with NahE (112 nM) over 60 min. **C)** *E. coli* crude lysate with expressed PhdJ and HBP (50 µM) at 0 min (red), 58 min (green), and purple (20 h). **D)** HBP (160 µM) and PhdG (1.1 µM) at 0 (blue), 14 (red), and 23 min (green).

Non-enzymatic Isomerization of HBP and HCCA. HBP (75 µM) was converted to HCCA (**Figure 2-2A**) at a rate of 2.2 ± 0.6 nM/s, giving a rate constant of 2.9 x 10^{-5} s^{-1}. HCCA (75 µM) was converted to HBP at a rate of 3.4 ± 0.2 nM/sec giving a rate constant 2.4 ± 0.3 x 10^{-5} s^{-1}.

Absorbance of HBP basic pH. An absorbance peak is observed near 400 nm under alkaline conditions indicating deprotonation of the *o*-hydroxyl substituent (**Figure 2-2B**).

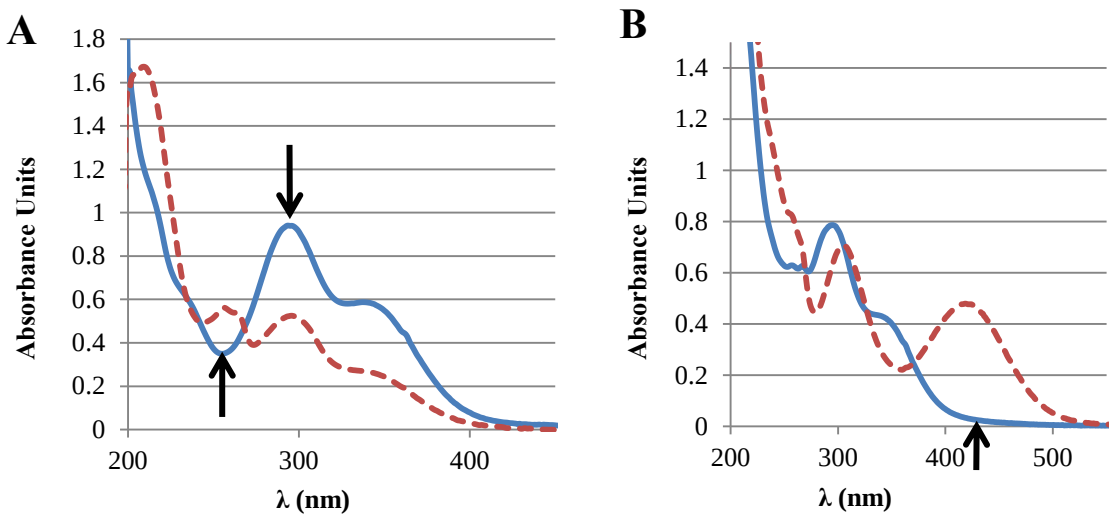


Figure 2-2 Isomerization of HBP over time and at basic pH.

A) HBP absorbance when dissolved in 50 mM sodium phosphate buffer (pH 7.0) (blue, solid) and after 20 h (red, dashed) **B)** Absorbance of HBP dissolved in 50 mM sodium phosphate buffer at pH 7.0 (blue, solid) and at pH 12.5 (red, dashed).

NahE was inhibited by salicylaldehyde (**Figure 2-3A**) with an IC_{50} near 20 μM. Kinetic data at various HBP and salicyaldehyde concentrations show the inhibition is not competitive (**Figure 2-3B**). Increasing HBP concentrations do not overcome inhibition by salicyaldehyde, suggesting salicylaldehyde inhibits the enzyme at a potential allosteric site, as data for the *Pseudomonas vesicularis* DSM 6383 homologue suggests (Kuhm, *et al.*, 1993).

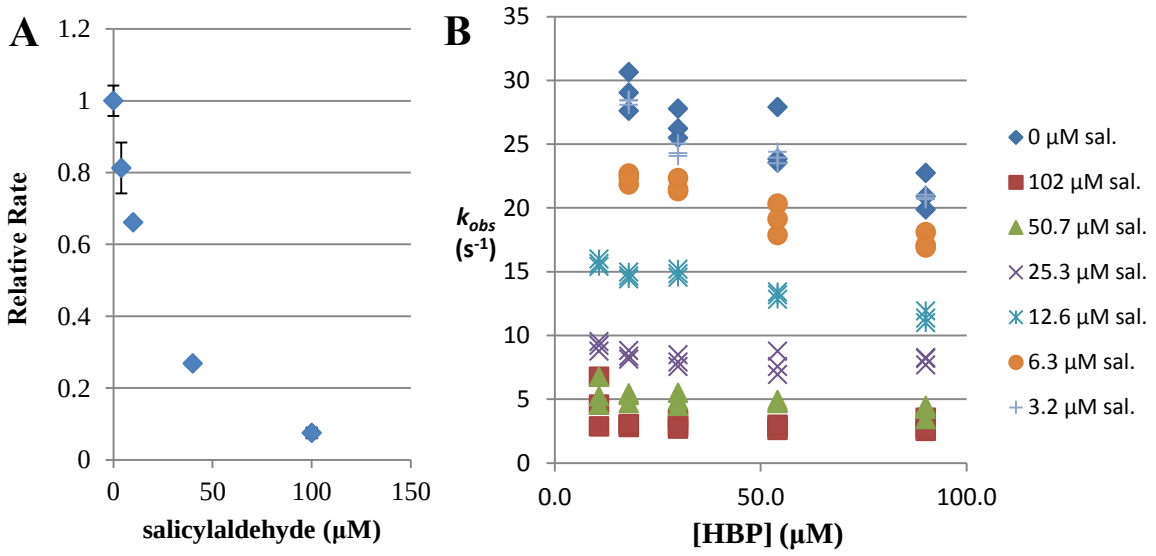


Figure 2-3 A) Initial rates of NahE using HBP (98 µM) with salicyaldehyde in 50 mM sodium phosphate buffer (pH 7.0). **B)** Rates obtained with variable HBP and salicylaldehyde (procedure is in **Section 2.2.3**).

When K183A NahE (11 µM) was added to 50 mM sodium phosphate buffer (pH 7.0) containing HBP (~100 µM), the observed rate was 2.2% that of the HBP reaction with 110 nM wild-type NahE. Additionally, the K183A mutant was not modified by HBP or 2-OP in the presence of $NaCNBH_3$ (**Table 2-5**), whereas wild-type NahE was modified. Taken together, these data confirm the importance of Lys-183 in covalent catalysis by NahE. CBP was converted to CBA and pyruvate by both NahE and PhdJ, but not PhdG (**Figure 2-4**), whereas BP and *trans*-cinnamaldehyde (with pyruvate) were converted to 2-oxo-6-phenyl-3,5-hexadienoate by all three enzymes (conversion by PhdJ and NahE has not been confirmed by NMR, **Figure 2-5 and 2-6**). *trans*-4-

Nitrocinnamaldehyde (with pyruvate) was converted by PhdG to 2-oxo-6-phenyl[4-nitro]-3,5-hexadienoate (**Figure 2-6D**).

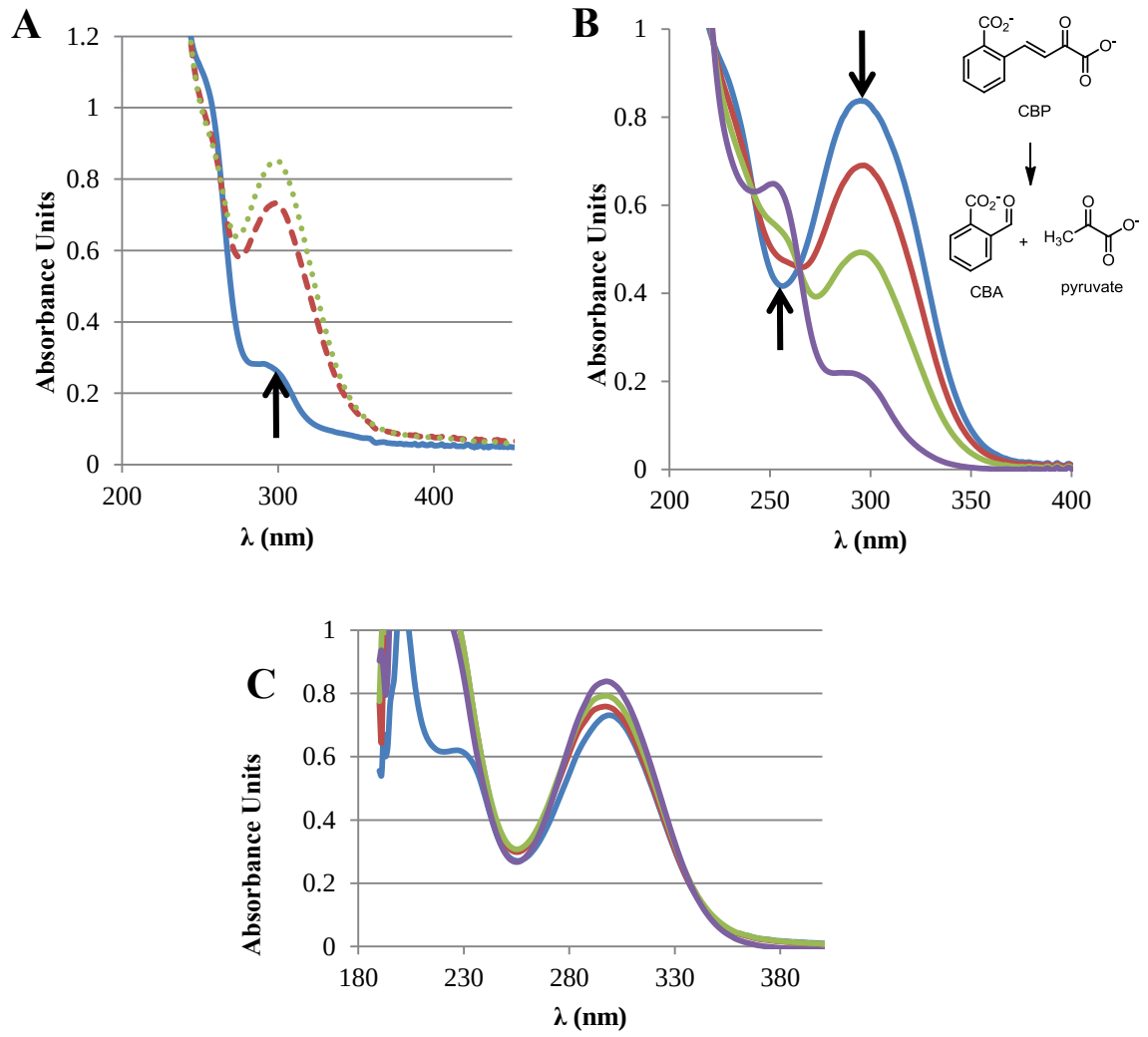


Figure 2-4 Spectra showing the reaction of CBP in the presence of NahE, PhdJ, and PhdG.

A) Incubation of CBA (50 µM) and pyruvate (2.5 mM) with supernatant of *E. coli* lysate containing expressed NahE after 0 h (solid, blue), 2.5 h (red, dashed), and 25 h (green, dotted). **B)** CBP (128 µM) and 6.5 µg/mL PhdJ (180 nM) after 0 min (blue), 1 min (red), 2 min (green), and 11 min (purple). **C)** CBP (87 µM, blue) alone and with PhdG (1.1 µM) after 0 min (red), 1 min (green), 364 min (purple).

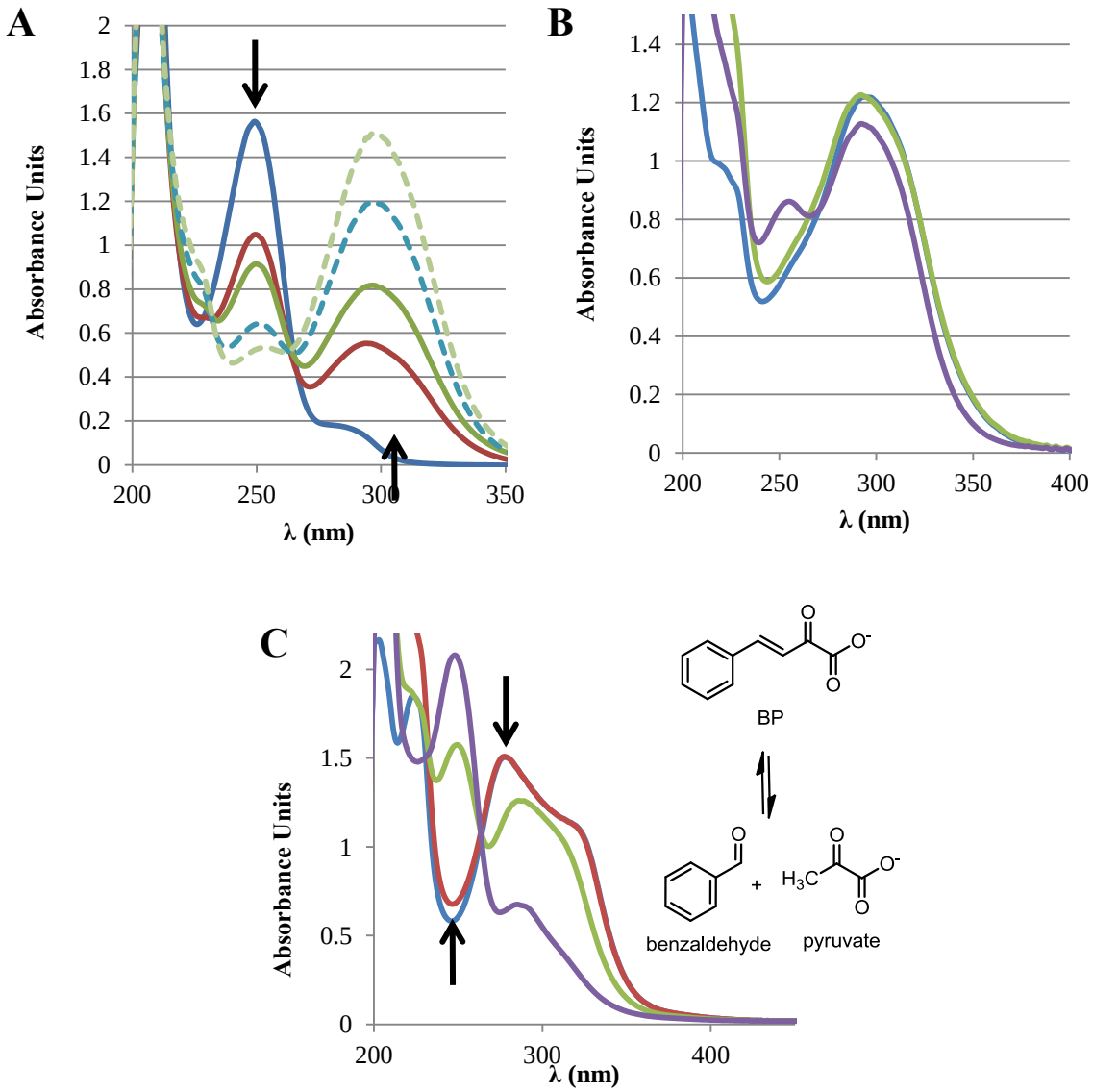


Figure 2-5 BP production/consumption by NahE, PhdJ, and PhdG

A) NahE (39 nM) with benzaldehyde (152 µM) and pyruvate (79 µM) in 100 mM dibasic sodium phosphate buffer (pH ~8) after 0 min (blue), 3 min (red), 7 min (green), 21 min (blue, dashed), 100 min (green, dashed) **B)** BP (30 µM) and PhdJ (90 nM) after 0 min (blue), 5 min (green), and 122 min (purple). **C)** BP (blue) and PhdG (1.1 µM) after 0 min (red), 3 min (green), and 14 min (purple).

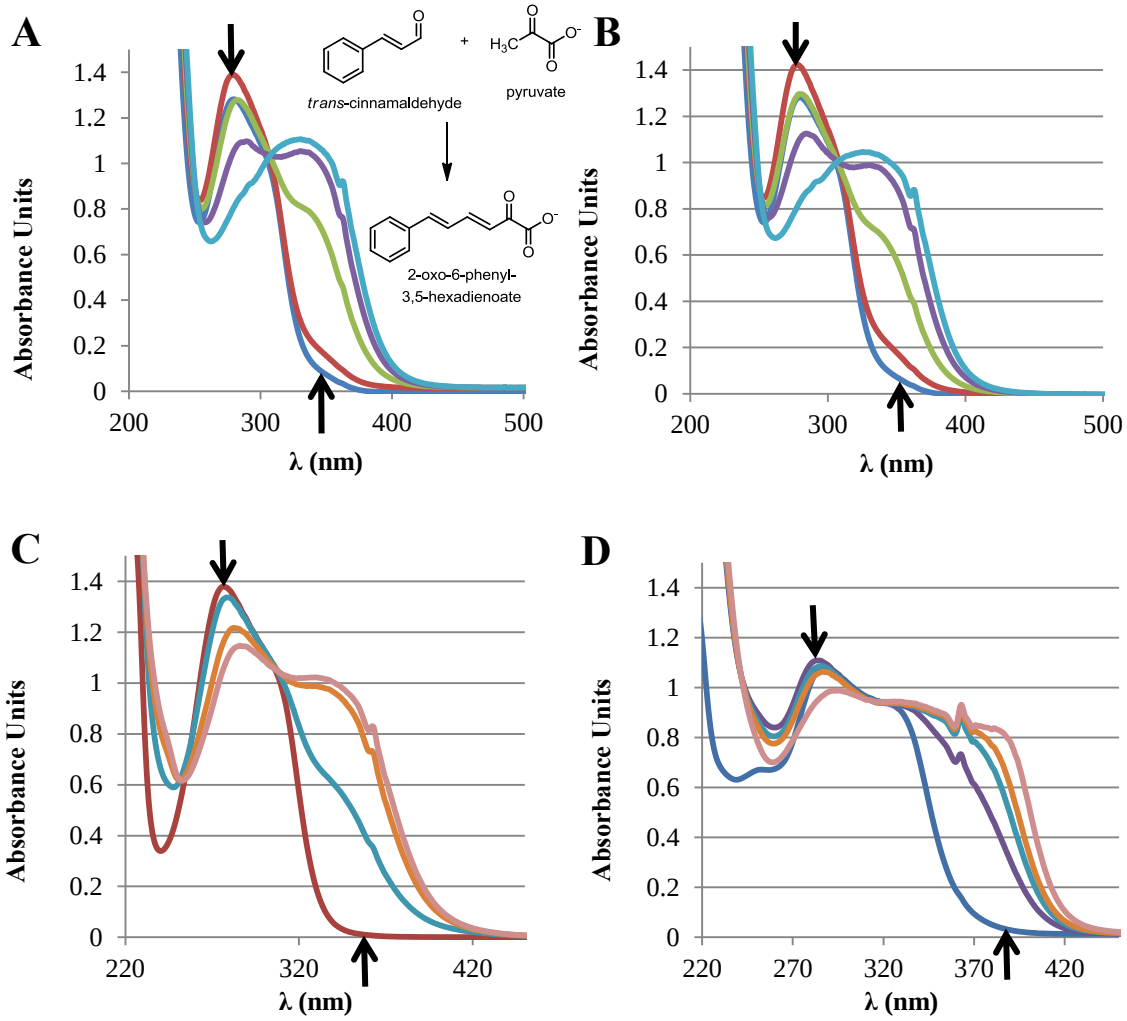


Figure 2-6 Absorbance spectra during conversion of *trans*-cinnamaldehyde and pyruvate to 2-oxo-6-phenyl-3,5-hexadienoate by NahE, PhdJ, and PhdG.

A) *trans*-cinnamaldehyde (76 µM) and pyruvate (5.9 mM, blue) with NahE (3.5 µM) after 0 min (red), 3 min (green), 11 min (purple), 30 min (light blue). **B)** Same as A, except PhdJ (0.69 µM) replaces NahE **C)** *trans*-cinnamaldehyde (100 µM) and pyruvate (1 mM) (red) with PhdG (1.1 µM) after 0 min (blue), 1 min (orange), 17 min (pink). **D)** *trans*-4-nitrocinnamaldehyde (~100 µM) and pyruvate (1 mM, blue) with PhdG (1.1 µM) after 0 min (purple and light blue), 1 min (orange), and 9 min (pink).

NahE, PhdJ, and PhdG did not process *trans*-cinnamaldehyde, 2-OP, or 4-phenyl-2-oxo-3-butynoate (POB). NahE did not convert HCCA. Additionally, none of the enzymes processed 1-hydroxy-2-naphthaldehye or 2-hydroxy-1-naphthaldehyde in the presence of pyruvate.

2.3.2 Steady-State Kinetic Parameters for NahE, PhdJ, PhdJ Mutants, and PhdG

Steady State Kinetic Parameters for NahE Using HBP. The K_M value for the NahE-catalyzed conversion of HBP to salicylaldehyde and pyruvate is low (1.2 μM), and was difficult to measure because of the low absorbance near 1 μM. Coupled assays using lactate dehydrogenase and salicyaldehyde dehydrogenase failed to give consistent rates at low concentrations. Eventually, an uncoupled assay protocol at low NahE concentration (3.2 nM) and at extended time courses allowed rate determination near 1 μM HBP. The NahE-catalyzed reaction was also affected by substrate inhibition (**Figure 2-7**) as HBP inhibited NahE with a K_I value of 61 ± 9 μM. It is unclear whether the inhibition is competitive, where HBP binds distal from the active site (as determined for salicylaldehyde) and allosterically decreases activity, or if inhibition is uncompetitive, where HBP binds to the pyruvate Schiff base complex, for example, and prevents regeneration of the active site. It is possible NahE could be affected by mixed inhibition, where both types of inhibition occur simultaneously.

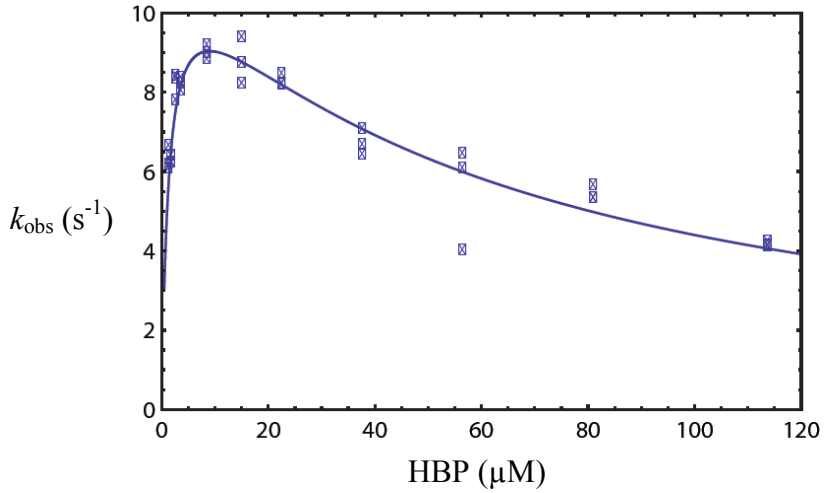


Figure 2-7 Initial rates for NahE with HBP reveal the enzyme is inhibited by substrate and possesses a low K_M value (20-fold less than K_I).

Steady state kinetic parameters for these reactions are shown in **Table 2-1**. These values show NahE prefers HBP as a substrate (vs. CBP or BP). The high k_{cat}/K_M value for NahE using HBP (1×10^7 $M^{-1}s^{-1}$) is partly due to the low K_M value that is ~100-fold less than for CBP and BP (in their conversion to CBA and benzaldehyde, respectively, and pyruvate). The k_{cat}/K_M value for NahE using HBP is ~10^4-fold higher than that using CBP or BP. The k_{cat} value for NahE using HBP is 45-fold higher than NahE using CBP and 120-fold higher than NahE using BP. The k_{cat}/K_M for PhdJ with CBP is 26-fold higher than PhdJ using HBP, mostly due to the higher k_{cat}, and 1000-fold greater than PhdJ using BP.

The k_{cat}/K_M value using HBP is 3200-fold higher for NahE than that for PhdJ. The difference is due to both k_{cat} and K_M, as the k_{cat} for NahE is 42-fold higher than for PhdJ,

and the K_M value for NahE is 75-fold lower than for PhdJ. The k_{cat}/K_M value for PhdJ using CBP is 68-fold higher than the k_{cat}/K_M for NahE using CBP, due to both k_{cat} and K_M. The k_{cat} for NahE is decreased 25-fold relative to PhdJ, and the K_M increased by 2.6-fold. The k_{cat}/K_M for NahE using BP is 7.6-fold greater than for PhdJ.

Table 2-1 Steady-state parameters of NahE and PhdJ with HBP, CBP and BP.

	NahE			PhdJ		
	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1}s^{-1}$)
HBP	11.7 ± 0.6	1.2 ± 0.2	$(1.0 \pm 0.2) \times 10^7$	0.28 ± 0.04	90 ± 20	$(3.1 \pm 0.8) \times 10^3$
CBP	0.26 ± 0.06	210 ± 60	$(1.2 \pm 0.5) \times 10^3$	6.5 ± 0.4	79 ± 7	$(8.2 \pm 0.9) \times 10^4$
BP	0.10 ± 0.01	160 ± 20	600 ± 100	Did Not Saturate		79 ± 1

Comparison of the S278N and D282E PhdJ Mutants Using HBP. Steady state kinetic parameters for the S278N and D282E PhdJ mutants are shown in **Table 2-2**. The K_M values for S278N and D282E mutants with HBP are 2-fold lower than for wild-type, and the k_{cat} values are reduced 54- and 35-fold, respectively. The S278N and D282E mutants were designed to the mimic residues of NahE interacting with the *o*-hydroxyl group of HBP. The parameters consistent with the residues being more significant in the binding of HBP than turnover. The k_{cat}/K_M value for the S278N and D282E PhdJ mutants using HBP is 24-fold and 16-fold less than wild-type. The k_{cat}/K_M value for NahE is ~10^4-fold greater than the values for S278N PhdJ and D282E PhdJ.

Comparison of the S278N and D282E PhdJ Mutants Using CBP. The k_{cat}/K_M value of the S278N and D282E PhdJ mutants for conversion of CBP to CBA and pyruvate are reduced 23-fold and 1500-fold, respectively, compared to wild-type. These values demonstrate the importance of Asp-282 in the PhdJ conversion of CBP (to CBA and pyruvate) relative to HBP or BP. The k_{cat}/K_M value for S278N PhdJ using CBP is 3-fold greater than that for NahE, but the k_{cat}/K_M for the D282E mutant using CBP is 23-fold less than that for NahE.

Comparison of the S278N and D282E PhdJ Mutant Using BP. The k_{cat}/K_M values for the S278N and D282E PhdJ mutants in their conversion of BP to benzaldehyde and pyruvate were 2.2-fold and 34-fold greater than that for wild-type PhdJ. These increases in k_{cat}/K_M values when arranging the PhdJ active site to be more NahE-like make the values using BP closer to those determined for wild-type NahE. The k_{cat}/K_M for S278N PhdJ is 3.5-fold less than that for NahE, and the value for D282E PhdJ is 4.5-fold greater than for NahE.

Table 2-2 Kinetic parameters of the S278N and D282E PhdJ mutant using HBP, CBP, and BP.

	PhdJ S278N			PhdJ D282E		
	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1}s^{-1}$)
HBP	$(5.2 \pm 0.4) \times 10^{-3}$	39 ± 6	130 ± 40	$(8 \pm 1) \times 10^{-3}$	40 ± 10	200 ± 60
CBP	Not Detectable		$(3.6 \pm 0.3) \times 10^{3}$	Not Detectable		53 ± 7
BP	Not Detectable		170 ± 2	Not Detectable		$(2.7 \pm 0.1) \times 10^{3}$

Comparison of Wild-type Enzymes Using BP. The k_{cat} value for the PhdG conversion of BP to benzaldehyde and pyruvate is ~4.5-fold greater than for NahE, where the K_M is about three-fold lower, giving a k_{cat}/K_M that is 15-fold higher. The k_{cat}/K_M value for PhdG using BP is 110-fold greater than with PhdJ.

Table 2-3 Steady-state parameters of PhdG using BP.

	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
BP	0.44 ± 0.08	50 ± 10	$(9 \pm 2) \times 10^3$

2.3.3 Steady-State Kinetic Parameters of NahE and PhdG Using Pyruvate and an Aldehyde

Initial rates for NahE- and PhdG-catalyzed conversion of salicyaldehyde and cinnamaldehyde, respectively, with pyruvate to form HBP and 2-oxo-6-phenyl-3,5-hexadienoate, respectively, were plotted with respect to both substrates (**Figure 2-8A,B**), and fit to **eq. 2** (**Section 2.2.3**), used for ping-pong mechanisms. The kinetic parameters for NahE and PhdJ using salicylaldehyde and cinnamaldehyde, and pyruvate, are summarized in **Table 2-4**.

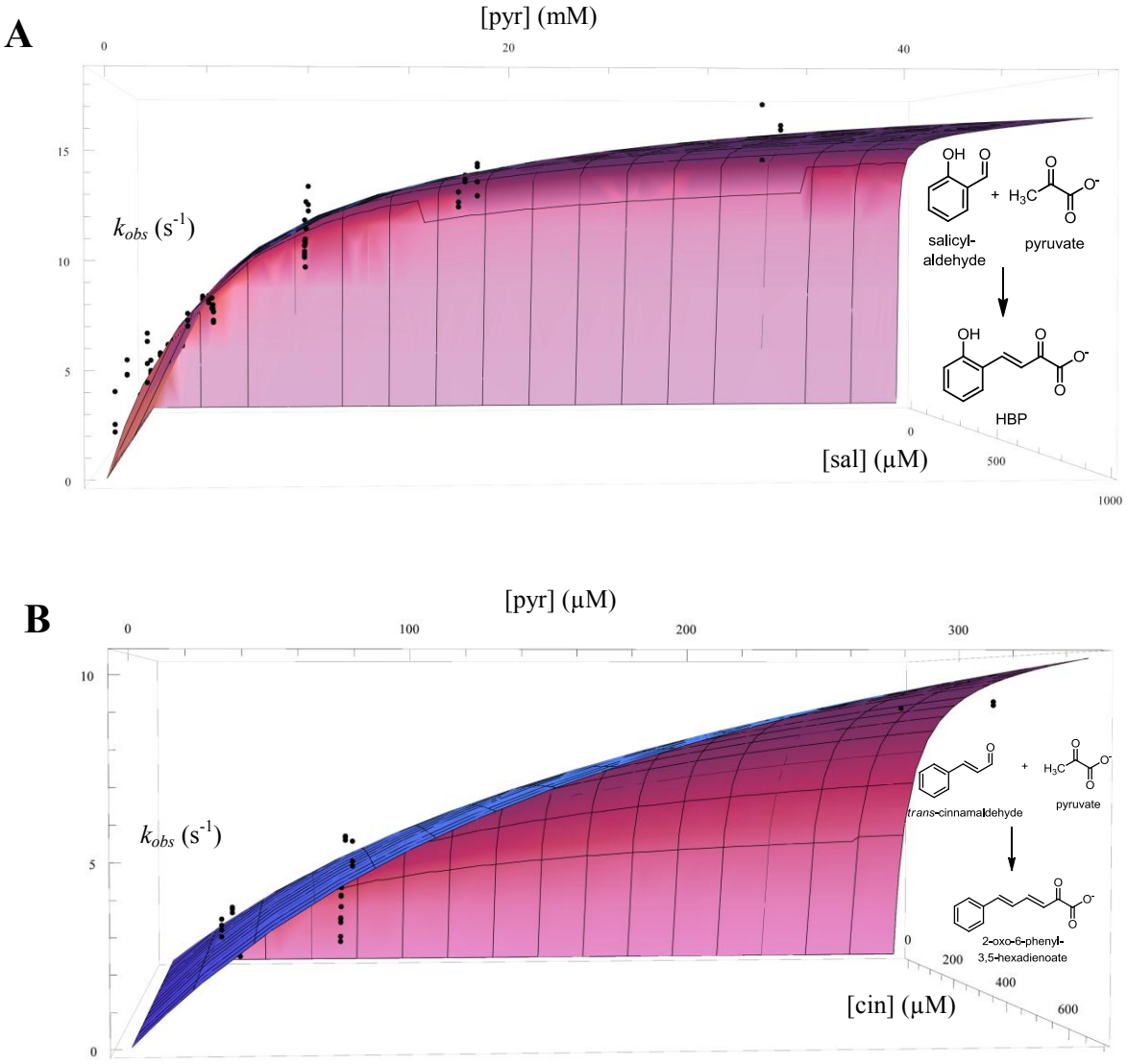


Figure 2-8 3-D plot of k_{obs} vs. conc. and fits to bisubstrate reactions.

A) NahE using salicyaldehyde (sal) and pyruvate (pyr) and **B)** PhdG using cinnamaldehyde (cin) and pyruvate.

Comparison of NahE- and PhdJ-catalyzed Aldol Condensation Reactions. The affinity of hydratase-aldolases and pyruvate varies in the low mM range. $K_{M,pyr}$ for PhdG

is 22-fold less than that for NahE. For NahE, the $K_{M,pyr}$ value is 750-fold greater than $K_{M,ald}$. For PhdG, the $K_{M,pyr}$ value is 2.3-fold greater than that for $K_{M,\ ald}$.

Table 2-4 NahE and PhdG kinetic constants with aldehyde (ald, specifically salicylaldehyde and cinnamaldehyde, respectively) and pyruvate.

Enzyme	k_{cat} (s^{-1})	$K_{M,\ ald.}$ (μM)	$K_{M,\ pyr.}$ (μM)	$k_{cat}/K_{M,\ ald}$ ($M^{-1}s^{-1}$)	$k_{cat}/K_{M,\ pyr}$ ($M^{-1}s^{-1}$)
NahE	18.7 ± 0.5	8 ± 1	(6.0 ± 0.4) × 10^3	(2.3 ± 0.3) × 10^6	(3.1 ± 0.2) × 10^3
PhdG	20 ± 2	120 ± 20	270 ± 50	(1.7 ± 0.3) × 10^5	(7 ± 2) × 10^4

2.3.4 ^{1}H NMR Spectrometric Analysis of Reactions Catalyzed by NahE, NahD, PhdJ, and PhdG

^{1}H NMR Analysis of NahE-catalyzed Reactions. ^{1}H NMR peak integrations for pyruvate (product) formation in three reactions were compared to the integrations of substrate peaks after 33 min. HCCA was converted to salicylaldehyde and pyruvate and was 9.7 % complete after 33 min (**Figure 2-9A**) in the presence of NahE. HBP was converted to salicylaldehyde and pyruvate and 16% complete at 33 min (**Figure 2-9B**) in the presence of NahE. A mixture of HCCA/HBP was converted fastest to salicylaldehyde and pyruvate (26% complete after 33 min, **Figure 2-9C**) in the presence of isomerase NahD and NaheE. An aldehyde peak is barely visible in a spectrum of the nonenzymatic control which was recorded after an overnight incubation, showing that the non-enzymatic rate is slow compared to reactions with NahE. NMR was used to confirm catalysis in the reverse direction, and for NahE conversion of benzaldehyde and pyruvate

to BP. ^{1}H NMR chemical shifts for HBP and HCCA are reported in **Section 2.2.2**. **salicylaldehyde:** ^{1}H NMR (H_2O, 600 MHz) δ 6.794 (1H, d, *J* = 9.0), 6.862 (1H, t, *J* = 7.47 Hz), 7.411 (1H, t, *J* = 7.77 Hz), 7.539 (1H, d, *J* = 7.74 Hz), 9.770 (1H, s) **pyruvate:** ^{1}H NMR (H_2O, 600 MHz) δ 2.168 (3H, s); **benzaldehyde:** ^{1}H NMR (H_2O, 600 MHz) δ 9.749 (1H, s), 7.778 (2H, d, *J* = 7.44 Hz), 7.577 (1H, t, *J* = 7.39 Hz), 7.443 (2H, t, *J* = 7.43)

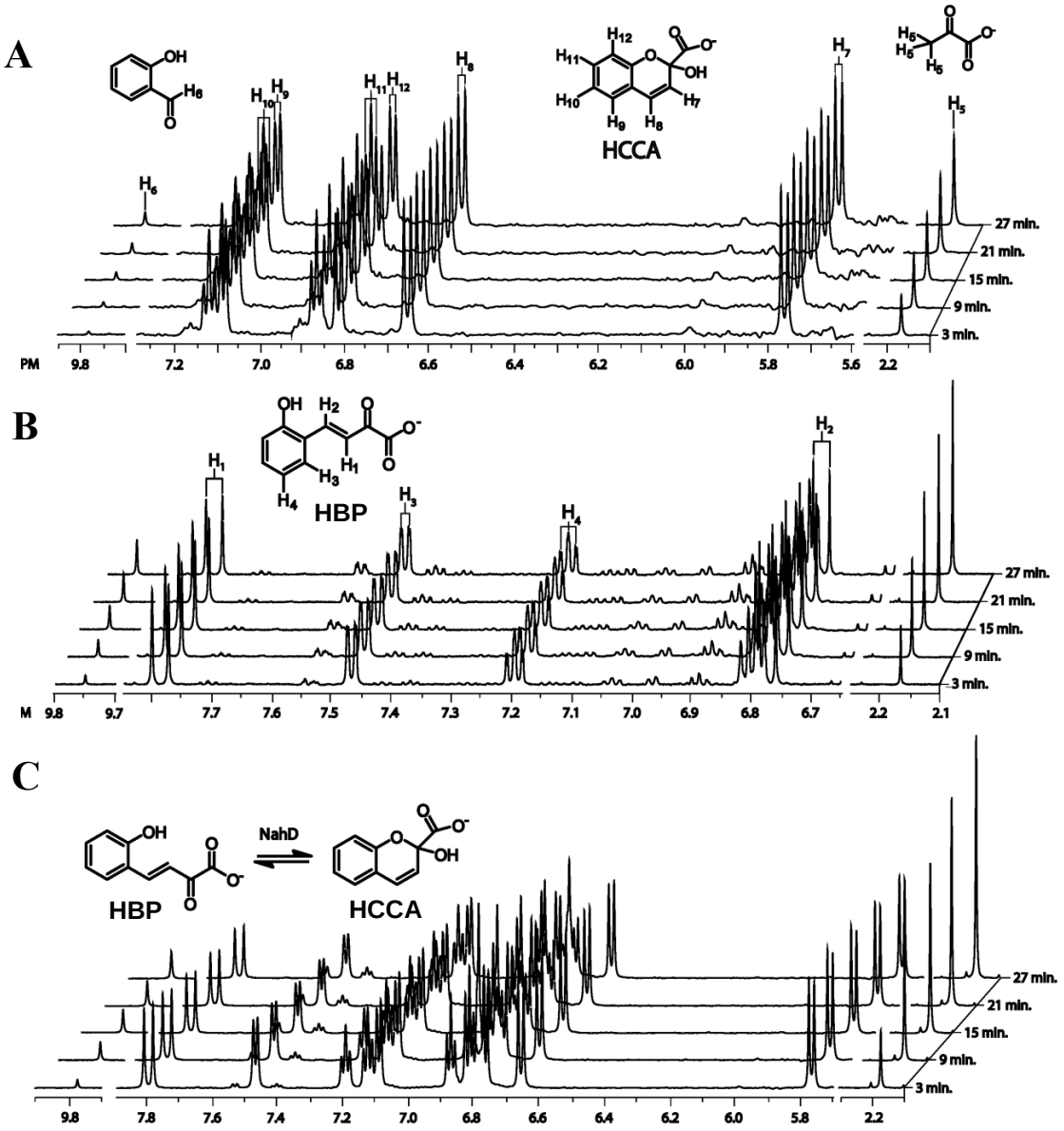


Figure 2-9 ^{1}H NMR spectra of NahE and NahD using HBP and HCCA.

A) NahE (16 μM) and HCCA (30 mM) **B)** NahE (16 μM) and HBP (30 mM) **C)** HCCA/HBP equilibrium (30 mM), NahE (16 μM) and NahD (~5 μM).

^{1}H NMR Analysis of PhdJ-catalyzed Reactions. ^{1}H NMR spectra following CBP with PhdJ indicated that CBP was consumed and CBA and pyruvate were produced over the course of 33 min. NMR was used to confirm catalysis in the reverse direction. ^{1}H NMR chemical shifts for CBP are reported in **Section 2.2.2**. CBA: ^{1}H NMR (H_2O, 600 MHz) δ 9.99 (1H, s), 7.68 (1H, d, *J* = 7.694 Hz), 7.5196 (1H, t, *J* = 8.002 Hz), 7.3862 (1H, d, *J* = 7.694 Hz);

^{1}H NMR Analysis of PhdG-catalyzed Reactions. ^{1}H NMR spectra indicated BP was processed by PhdG to form benzaldehyde and pyruvate (chemical shifts for benzaldehyde and BP are reported above and in **Section 2.2.2**, respectively) over the course of 33 min. *trans*-Cinnamaldehyde and pyruvate conversion by PhdG to 2-oxo-6-phenyl-3,5-hexadienoate was monitored by ^{1}H NMR over the course of 33 min. **2-oxo-6-phenyl-3,5-hexadienoate**: ^{1}H NMR (H_2O, 600 MHz) δ 6.20 (1H, d, *J* = 15.58 Hz), 6.92-7.03 (2H, m), 7.20-7.35 (4H, m), 7.196 (1H, td, *J* = 7.77 Hz, 1.65 Hz), 7.469 (1H, dd, *J* = 7.86 Hz, 1.62 Hz), 7.425 (1H, s), 7.438 (1H, s).

2.3.5 Modification of NahE, PhdJ, and PhdG by Various Aldehydes and α-Keto Acids

UV/Vis Spectra of NahE in the Presence of Salicylaldehyde, Benzaldehyde, and CBA. When salicylaldehyde (173 μM) was incubated with NahE (86.5 μM), the initial peak near 325 nm decreased and a peak emerged near 400 nm over the course of 90 min (**Figure 9-10**A). This indicates Schiff base formation between salicylaldehyde and NahE, similar to the absorbance spectra observed during Schiff base formation in PLP-

dependent enzymes (Osterman *et al.*, 1999). Similar absorbance changes were not observed when benzaldehyde or CBA (173 µM) were incubated with NahE (86.5 µM) for the same length of time (**Figures 9-10B,C**).

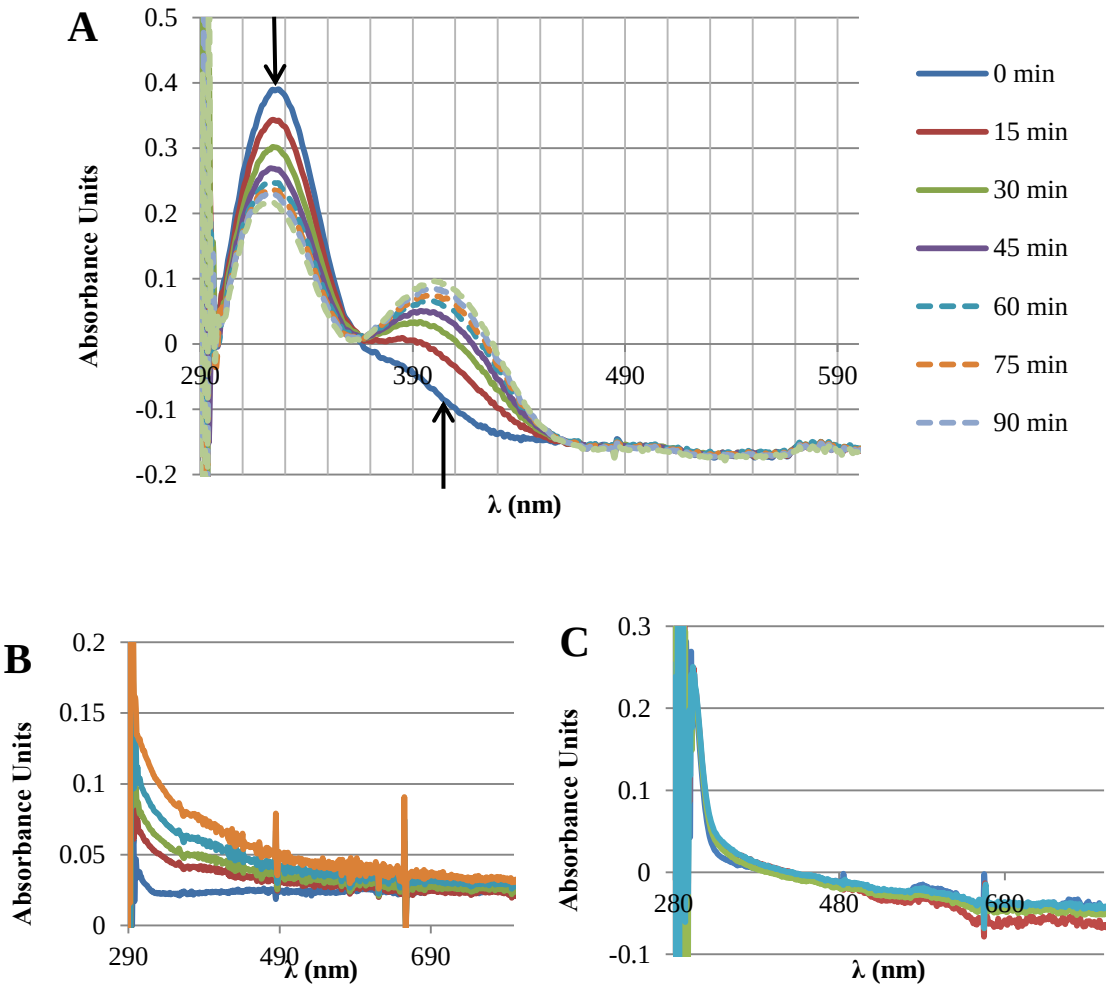


Figure 2-10 UV/Vis spectra of NahE and aldehyde substrates/products.

NahE (85.6 µM) in the presence of **A)** salicylaldehyde (173 µM); **B)** benzaldehyde (173 µM) at 0 min (blue), 15 min (red), 30 min (green), 60 min (light blue), and 120 min (orange); **C)** and CBA at 0 min (blue), 15 min (red), 30 min (green), and 44 min (light blue).

UV/Vis Spectra of PhdJ in the Presence of Salicylaldehyde, Benzaldehyde, and CBA. When salicylaldehyde (173 μM) was incubated with PhdJ (86.5 μM), the initial peak at 325 nM decreased and a new peak emerged near 400 nm over the course of 30 min, indicating a Schiff base had formed with salicylaldehyde and the enzyme (**Figure 9-11A**). Since the spectra at 15 min and 30 min are almost identical, equilibrium was reached in the first 15 min. The rate of Schiff base formation with salicyaldehyde was faster for PhdJ than that for NahE using the same concentration of salicylaldehyde and enzyme. Schiff base formation with benzaldehyde and CBA did not occur for PhdJ with the same amounts of enzyme and aldehyde during a similar time period (1-2 h) (**Figures 9-11B,C**).

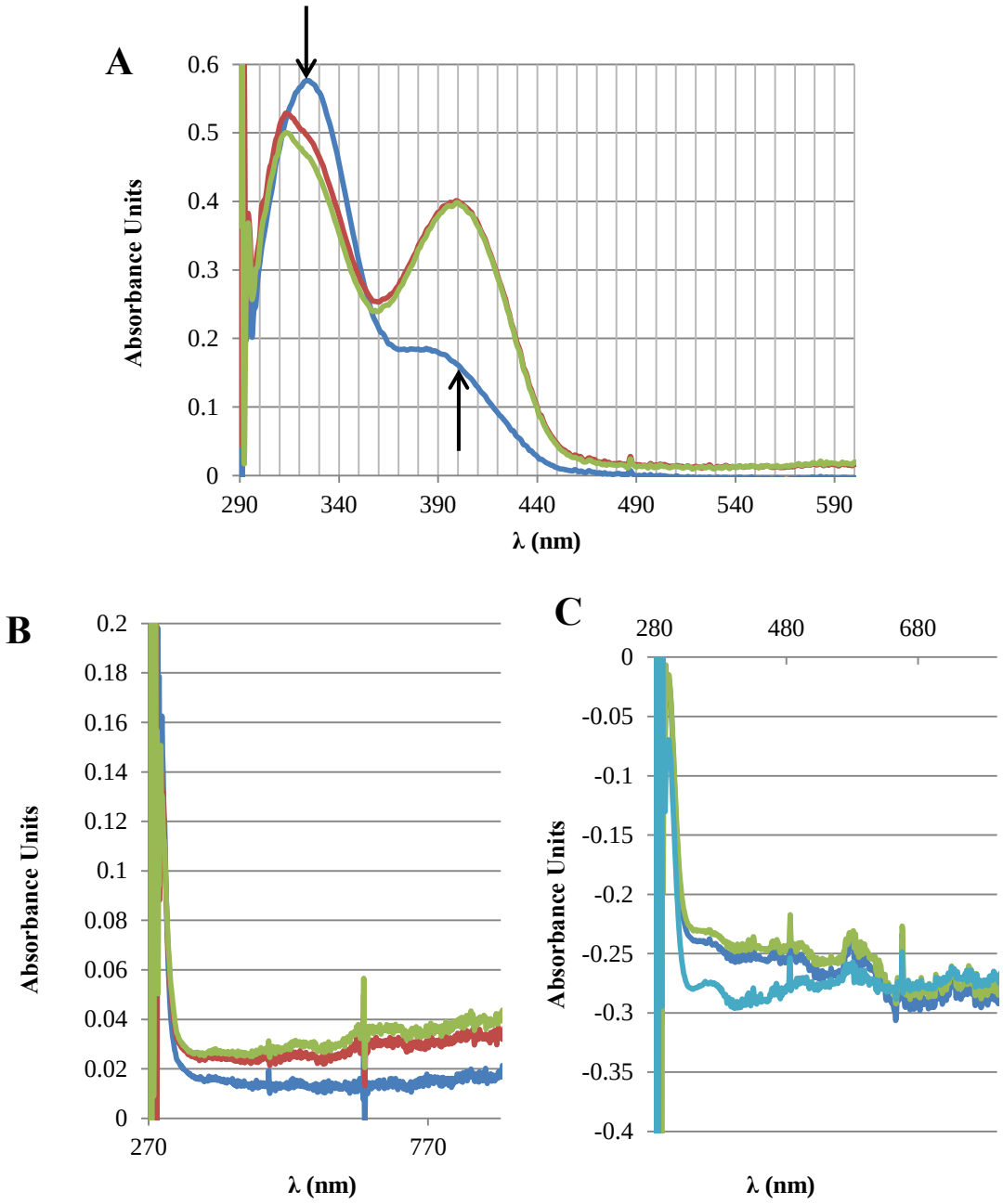


Figure 2-11 UV/Vis spectra of PhdJ and aldehyde substrates/products.

PhdJ (86.5 µM) with **A)** salicylaldehyde (173 µM), at 0 min (blue), 15 min (red), 30 min (green); **B)** benzaldehyde (173 µM) after 0 min (blue), 15 min (red), 30 min (green); **C)** CBA (173 µM) after 0 min (blue), 30 min (green), and 884 min (light blue).

Modification of NahE in the Presence of $NaCNBH_3$. NahE was incubated with HBP, salicylaldehyde, pyruvate, 2-OP, and POB in the presence of $NaCNBH_3$ and the residual activity was significantly decreased compared to the control sample with only $NaCNBH_3$ treatment (no other added compound). Mass increases in samples correspond to each added compound (except $NaCNBH_3$) minus one oxygen atom (within 5 Da), NahE had been modified by each compound with the loss of water (**Table 2-5**). Incubation of NahE with HBP resulted in mass increases corresponding to imine adducts with both HBP and pyruvate, showing HBP was occasionally processed before reduction of the Schiff base. Incubation of pyruvate with the enzyme resulted in mass increases indicative of labeling at one and two sites. Labeling with POB resulted in labeling at up to four sites. The observed mass of singly-labeled NahE indicated loss of an oxygen atom, whereas all other modifications occurred without this loss, suggesting Lys-183 forms a Schiff base with POB, followed Micheal addition to the acetylene compound by other nucleophilic amino acids of NahE.

Table 2-5 ESI-MS signal peak values with modification of NahE using NaCNBH3.

Modification of NahE in the presence of $NaCNBH_3$

modifier	calculated mass[a]	observed mass (unmodified)	observed mass (modified)	difference
None	36636.9	36641.0	not observed	-
HBP[b]	36813.1	36641.0	36818.0	177.0
sal	36743.0	36641	36744	103.0
pyr[c]	36709.0	not observed	36710	69.0
2-OP	36733.0	36641.0	36735.0	94.0
POB[d]	36809.1	not observed	36807.0	166.0

Table 2-6 ESI-MS signal peak values with modification of K183A NahE using NaCNBH3.

Modification of NahE K183A in the presence of $NaCNBH_3$

modifier	calculated mass	observed mass (unmodified)	observed mass (modified)	difference
No Treatment	36579.8	36585.0	not observed	-
HBP[e]	36756.0	36580.0	not observed	-
2-OP	36675.9	36585.0	not observed	-

[a]The calculated mass is without the N-terminal Met and accounts for the loss of an oxygen atom.

[b]A peak was observed at 36716 consistent with modification by pyruvate.

[c]A second modification by pyruvate was observed at 36783.

[d]Modification occurred with the loss of oxygen on first of four modifications. $NaCNBH_3$ was not used.

[e]The peak at 36614.0 (+34) is likely TFA contamination.

ESI MS/MS analysis found Lys-183 was labeled with pyruvate 10-fold more frequently than other lysine residues (**Figure 2-12**). Further evidence of Schiff Base formation occurs with Lys-183 comes from the observation that HBP and 2-OP did not label K183A NahE in the presence of $NaCNBH_3$ (**Table 2-6**).

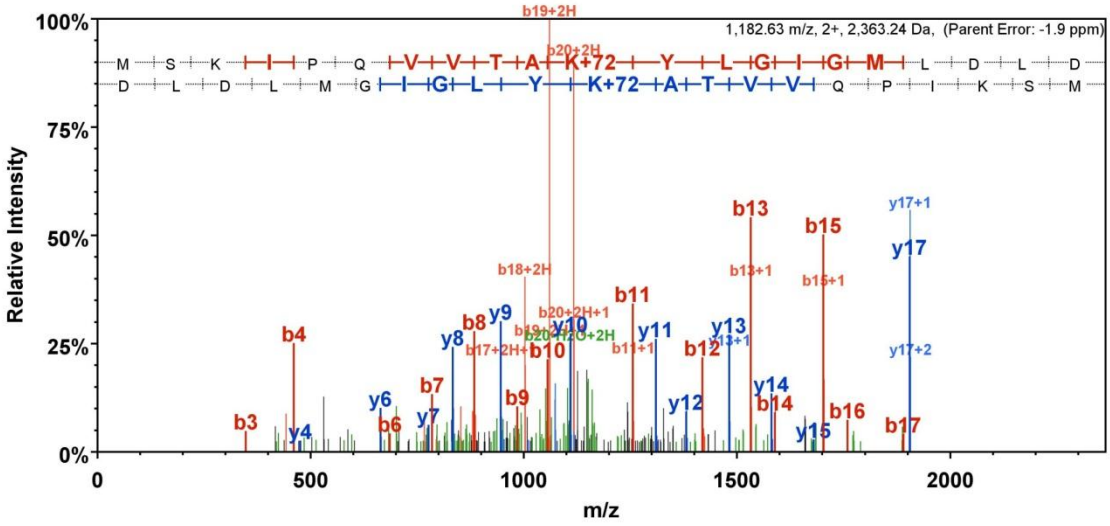


Figure 2-12 ESI-MS/MS spectrum of fragments from modified NahE incubated with pyruvate and $NaCNBH_3$.

PhdJ was incubated with CBP, CBA, and pyruvate in the presence of $NaCNBH_3$, and residual activity significantly decreased when compared to a control without added substrate or product with $NaCNBH_3$. Observed monomeric mass values (**Table 2-7**) indicated the labeling of these compounds occurred with oxygen loss, presumably through Schiff base formation with Lys-180.

Table 2-7 ESI-MS signal peak values with modification of PhdJ using CBP, CBA, and pyruvate in the presence of $NaCNBH_3$.

Modification of PhdJ in the presence of $NaCNBH_3$

modifier	calculated mass[a]	observed mass (unmodified)	observed mass (modified)	difference
No treatment	35966.8	35967	-	-
CBP[b]	36168.0	35967	36173	206.0
CBA[c]	36109.9	not observed	36102	
Pyruvate[d]	36037.9	35967	36039	72.0

[a]The calculated mass is without the N-terminal Met and accounts for the loss of an oxygen atom.

[b]Peaks observed at 36099 and 36233.

[c]Peaks observed at 36140, 36235, 36274, 36268, 36406, 36502, and 38640.

[d]A second modification by pyruvate was observed at 36112.

PhdG was incubated with 2-OP, POB, and pyruvate in the presence of $NaCNBH_3$ and residual activity significantly decreased compared to the control with $NaCNBH_3$ treatment only (no modifier). Observed monomeric mass values from ESI-MS data indicated these compounds modified PhdG with the loss of oxygen, likely via Schiff base formation with Lys-180 (**Table 2-8**). PhdG was not modified when incubated with CBP with $NaCNBH_3$.

Table 2-8 ESI-MS signal peak values with modification of PhdG using various compounds in the presence of $NaCNBH_3$.

Modification of PhdG in the presence of $NaCNBH_3$

modifier	calculated mass[a]	observed mass (unmodified)	observed mass (modified)	difference
No treatment	37454	37455	-	-
2-OP	37546	37454	37548	94
POB[b]	37628	37455	37623	168
Pyruvate	37627	37456	37528	72

[a]The calculated mass is without the N-terminal Met and accounts for the loss of an oxygen atom.

[b]Peaks observed at 37492 and 37651 may be a result of TFA contamination.

2.3.6 Dynamic Light Scattering

Two peaks were observed in the LC chromatograph (19.9 min and 22.4 min) for NahE. The first peak (97.5% mass fraction) had a molar mass moment of 145700 Da. This species is likely the tetrameric form of NahE (predicted MW: 146548 Da). The second peak (2.5% mass fraction) had a molar mass of 36660 Da and this species is presumed to be monomeric NahE, as its predicted monomeric mass is 36637. For PhdJ, four absorbance peaks in the LC chromatogram were observed (15.7, 16.9, 19.2, and 20.5 min). The second peak was the largest (88.0% mass fraction) with a molar mass moment of 138,300 Da. This molecular weight corresponds to the tetrameric state of PhdJ (predicted MW: 143995 Da). The fourth peak (4.8% mass fraction) had a molar mass moment of 421,500 Da and is consistent with trimerization of tetramers (predicted MW: 431956). The other two peaks that contained 7.2% of the mass of the sample are likely

due to impurities from the protein preparation. A single peak in the LC chromatogram for PhdG had a molar mass moment of 146700 Da, which corresponds to the tetrameric form of PhdG (predicted MW: 147160 Da).

2.3.7 Crystal Structures of NahE

The structure of NahE was solved to identify primary active site residues (those binding the pyruvate moiety) and secondary active site residues (not conserved within the NAL subgroup) and to determine their roles in substrate binding and turnover. The SeMet derivative was solved at 1.9 Å resolution by Multiwavelength Anomolous Dispersion (MAD), with two monomers per asymmetric unit. The structure is a tetramer (**Figure 2-13A**) composed of $(\alpha/\beta)_8$ barrels containing 3 C-terminal helices (**Figure 2-13B**). Crystal structures of NahE in complex with intermediates of the HBP- and BP-catalyzed reactions were obtained through soaking experiments and solved by molecular replacement using a Native NahE structure.

NahE and SeMet NahE formed rhombohedral crystals within a few days. The SeMet structure (**NahE-SeM**) was solved in the $P3_221$ space group (as were all NahE structure presented here) using MAD. NahE-SeM is structurally homologous to NAL and DHDPS (**Section 1.4**), as the monomeric RMSD value is 2.18 Å for NAL (PDB entry 1NAL) and 1.68 Å for DHDPS (PDB entry 1DHP). Significant deviations for NahE-SeM are seen in the loops and helices near the N-terminus of the β-barrel (**Figure 2-13C**). The active site of NahE-SeM shows Lys-183, Tyr-155, and Thr-65 and Phe-66 in the GXXGE motif, cores residues conserved in NAL subgroup members (**Figure 2-13D**). The active

site also contains secondary residues that are not conserved in the NAL family including Trp-128′, Asn-157, Asn-281, Trp-224, Glu-285, and Arg-288.

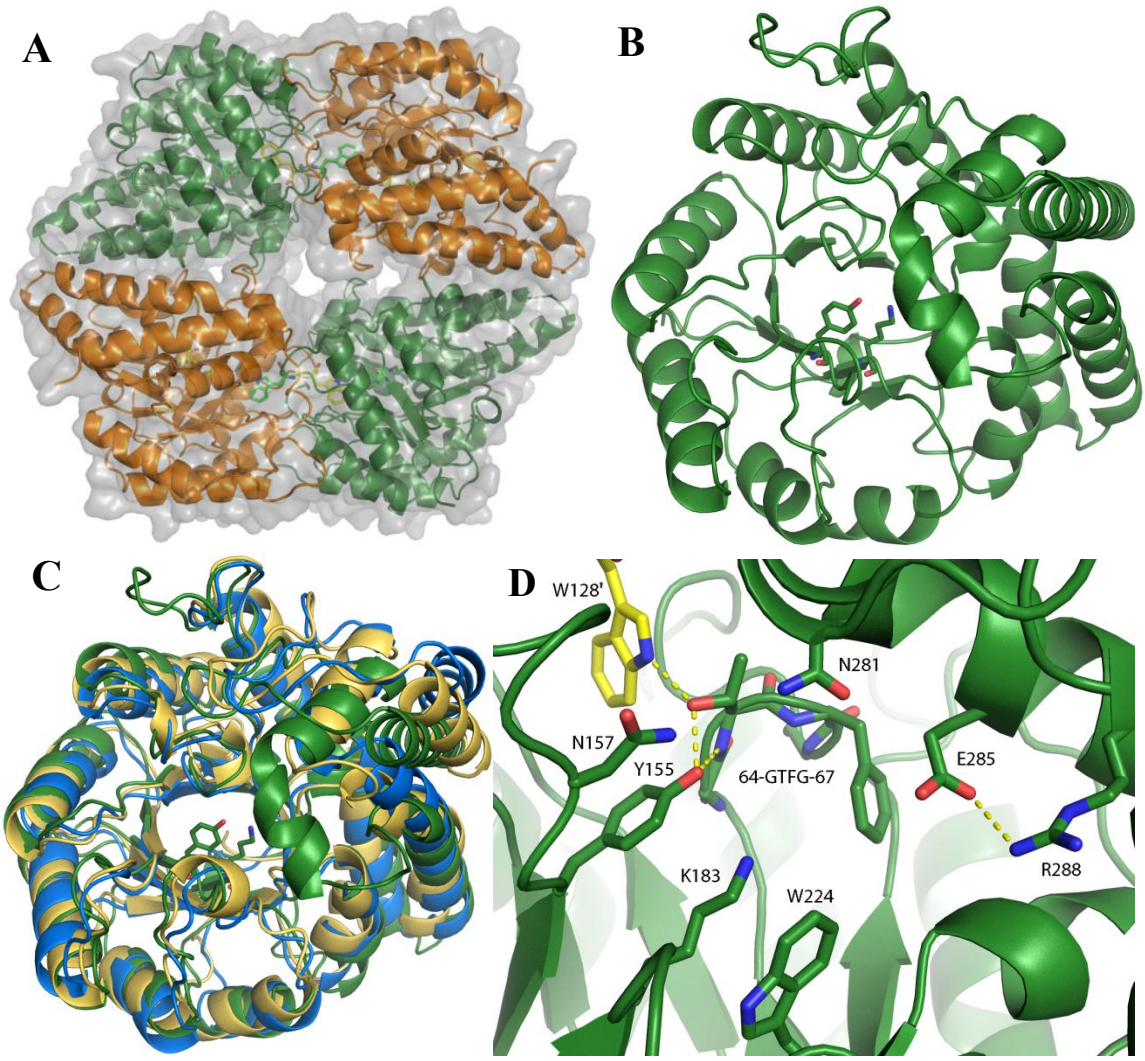
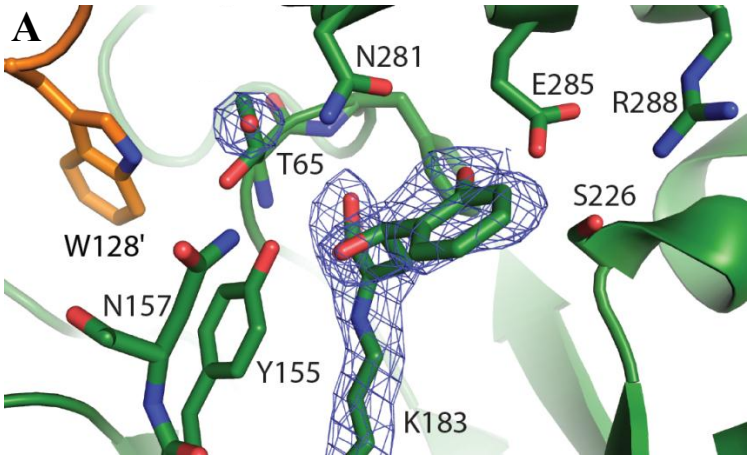
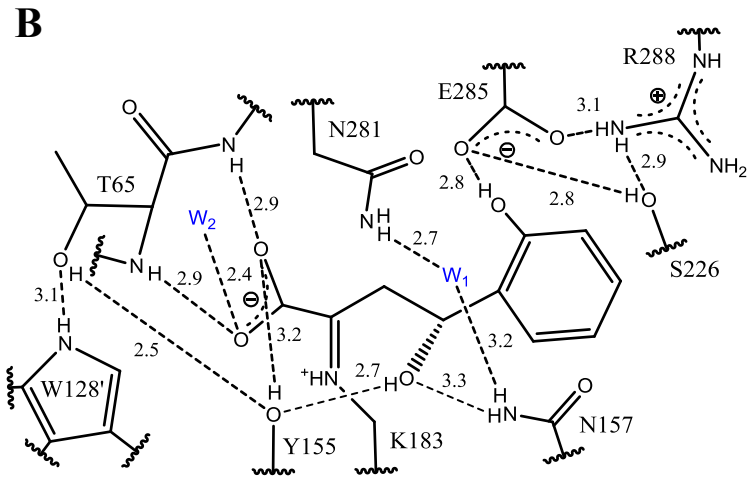


Figure 2-13 The 1.9 Å resolution crystal structure of NahE-SeM.

A) The NahE-SeM tetramer. **B)** NahE-SeM monomer with Tyr-155 and Lys-183 shown as sticks. **C)** Alignment of NahE-SeM (forest), NAL (marine, PDB entry 1NAL), and DHDPS (yellow orange, PDB entry 1DHP) **D)** Active site of NahE-SeM.

A crystal structure of NahE in complex with (*R*)-4-hydroxy-4-(2-hydroxyphenyl)-2-iminobutanoate (**HN**, for *o*-hydroxyl intermediate), the intermediate in the reaction with HBP was solved at 2.1 Å resolution by crystal soaking experiments with HBP. The ligand HN is bound to the enzyme via an imino linkage with Lys-183 (**Figure 2-14A,B)**. The hydroxyl group at C-4 is in the (*R*) configuration and forms hydrogen bonds with the side chain hydroxyl group of Tyr-155 and the side chain amide nitrogen of Asn-157. Interactions between Trp-128′, Tyr-155, the GXXGE motif, and the C-1 carboxylate group are like those observed for NAL, DHDPS, and D-2-keto-3-deoxygluconate aldolase (KDGA) (**Figure 2-14A,B**). The side chain amide nitrogens Asn-157 Nδ and Asn-281 donate a hydrogen bond to a water molecule near the C-4 hydroxyl group of the intermediate. The *o*-hydroxyl group on the aromatic ring of the intermediate forms hydrogen bonds with one side chain carboxylate of Glu-285. This carboxylate is stabilized by the side chains of Ser-226 and Arg-288 (Nη). An α-helix extends over the C-terminus of the β-barrel bringing Phe-269, Phe274, and Phe277 near Lys-183, potentially increasing the hydrophobicity of the active site (**Figure 2-14C**). The distance of the *o*-hydroxyl group and C-3 of the intermediate is 3.1 Å, which makes the group a possible proton donor to C-3 during the addition of water. Another candidate for the proton donor is the hydroxyl group of Tyr-155, which is 3.9 Å from C-3.


A
N281
E285
R288
T65
S226
W128'
N157
Y155
K183


B
R288
NH
E285
3.1
HN
H
2.9
NH2
2.8
2.8
S226
N281
T65
W2
2.9
2.7
W1
2.9
2.4
3.1
3.2
2.5
3.2
2.7
3.3
W128'
Y155
K183
N157

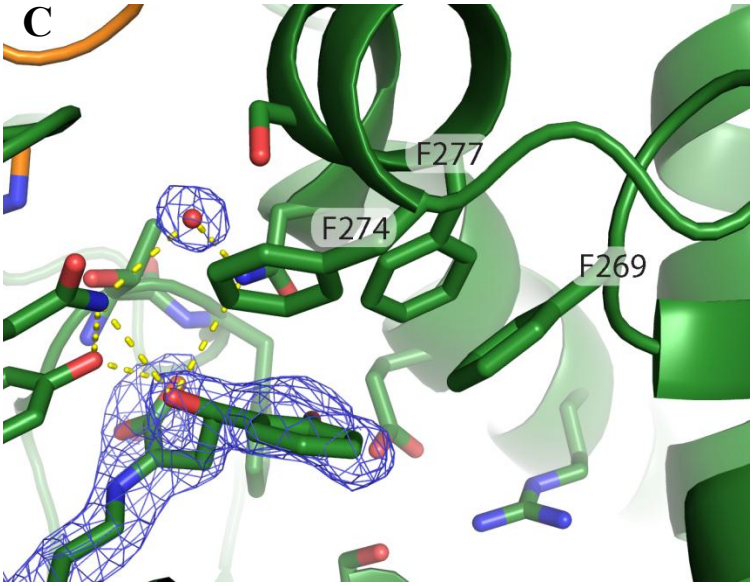

C
F277
F274
F269

Figure 2-14 Active site of the 2.1 Å resolution NahE-HN crystal structure.

A) Continuous $2F_o$-F_c electron density around Lys-183 and HN, contoured at 1.0 σ. **B)** Diagram of interactions near HN in the active site of NahE-HN. Distances are in Å. **C)** Phenylalanine residues near HN.

Another crystal structure of NahE (2.3 Å resolution) obtained in soaking experiments using HBP showed HN in one active site and a pyruvoyl adduct and salicyaldehyde in another active site (**NahE-SP**). A comparison of this structure, NahE-SP, to NahE-HN, shows similar interactions with the pyruvoyl moiety **(Figure 2-15A,B**). Additionally, the hydrogen bond formed between the *o*-hydroxyl group of intermediate HN and the side chain carboxylate of Glu-285 in NahE-HN is likely the one formed for the hydroxyl group of salicyaldehyde NahE-Sal. The side chain amide nitrogen of Asn-157 is 3.4 Å from the side chain hydroxyl group of Tyr-155, similar to the 3.3 Å distance in NahE-HN, but there is no longer a water molecule bridging the side chain amide nitrogen of Asn-157 (Nδ) and Asn-281 (Nδ). Interactions in NahE-SP between the *o*-hydroxyl group of salicylaldehyde, and side chains of Ser-226, Glu-285, and Arg-288 (Nη) are resembles those seen in NahE-HN.

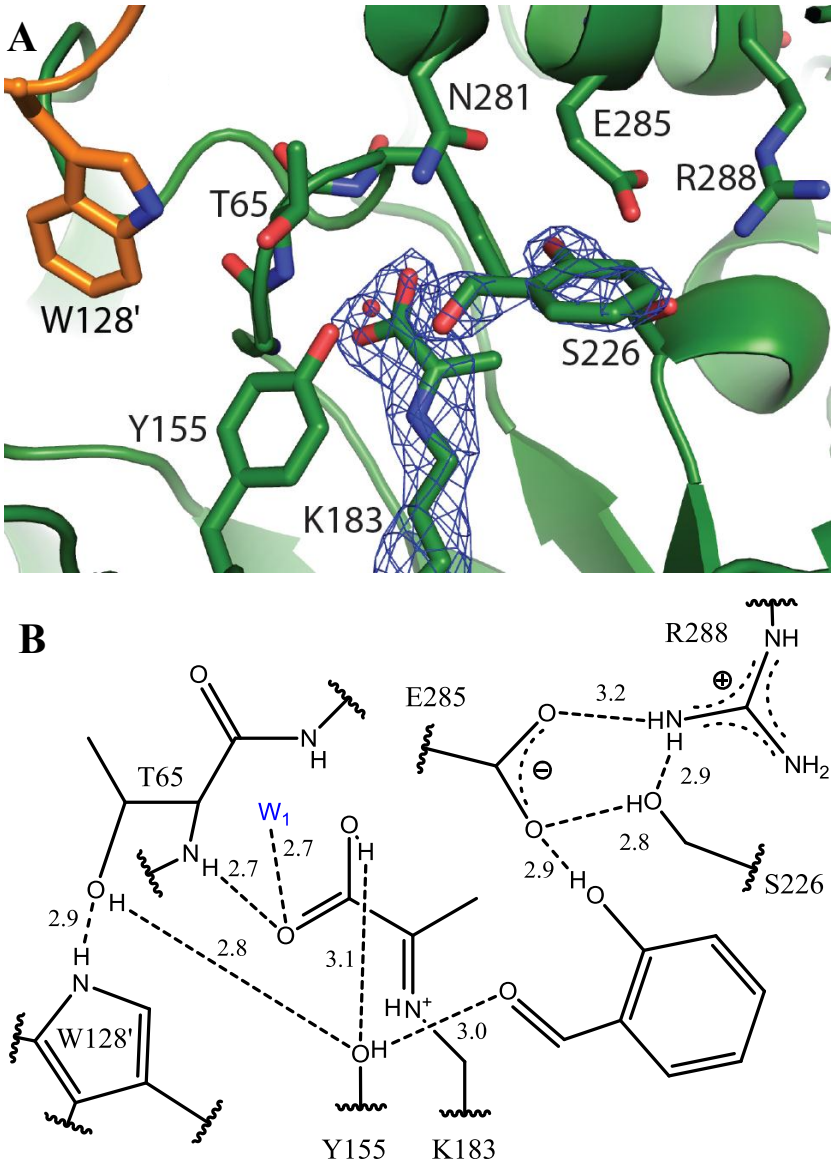


Figure 2-15 Active site of the 2.3 Å resolution crystal structure of NahE-SP.

A) $2F_o$-F_c electron density map, countered at 1.0 σ, surrounding Lys-183, the pyruvoyl adduct, and the product salicyaldehyde. **B)** Map of interactions in an active site of NahE-SP containing salicylaldehyde and pyruvate imine.

Data Collection	NahE-SeM	NahE-HN	NahE-SP
space group	$P3_221$	$P3_221$	$P3_221$
cell dimensions			
a, b, c (Å)	85.31, 85.31, 133.31	88.00, 88.00, 141.60	84.97, 84.97, 132.72
α, β, γ (deg)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
resolution (Å)	1.94-40.00	2.10 - 50.00	2.30-50.00
R_{sym} (%) (linear R factor)	0.150 (0.506)	0.129 (0.854)	0.128 (0.335)
I/σ	15 (4.5)	21 (2.3)	21 (4.47)
completeness (%)	100.0 (100.0)	100.0 (100.0)	97.7 (53.8)
Refinement			
resolution (Å)	38.09-1.94	44.01-2.10	49.30-2.32
no. of reflections	42138	37513	24619
R_{work}/R_{free} (%)	13.55/18.63	20.09/22.88	15.74/21.03
no. of atoms			
protein	5053	5496	5032
water	650	439	324
Average *B* factor (Å^2)			
protein	14.1	33.1	23.4
water	26.6	35.9	28.7
rmsd			
bond lengths (Å)	0.010	0.005	0.006
bond angles (deg)	1.140	0.869	1.032
Ramachandran plot (%)			
favored regions	98.3	98.1	98.0
allowed regions	1.7	1.9	2.0
disallowed regions	0.0	0.0	0.0

Data for the last resolution shell are given in parentheses.

Table 2-9 Crystallographic data and refinement statistics for NahE-SeM, NahE-HN, and NahE-SP.

A crystal structure of NahE (2.9 Å resolution) covalently linked to salicyaldehyde via Lys-183 (NahE-Sal) was obtained by soaking experiments. The *o*-hydroxyl group of

the covalent adduct likely hydrogen bonds (3.1 Å) with the backbone amide proton of Thr-65 (**Figure 2-16**). Due to the low resolution of NahE-Sal, it is possible that the *o*-hydroxyl is pointing in the opposite direction.

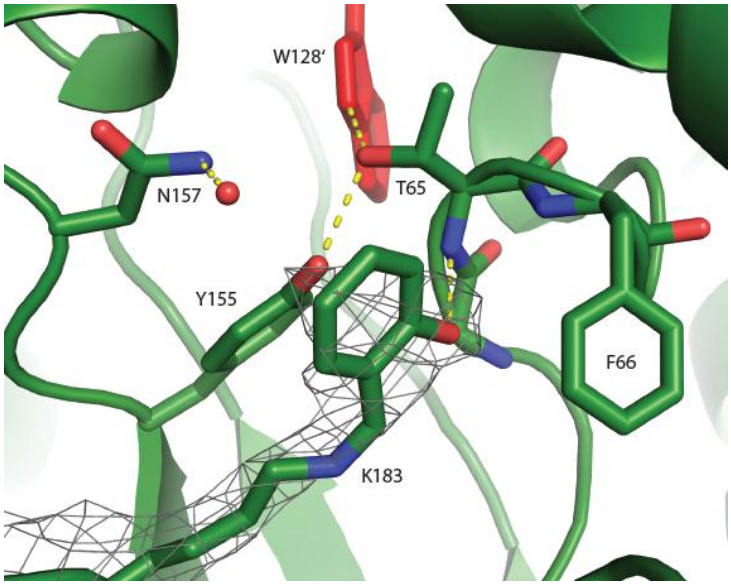


Figure 2-16 Active site of the 2.9 Å resolution crystal structure of NahE-Sal. $2F_o$-F_c electron density map surrounds Lys-183 and salicyaldehyde, contoured at 1.5 σ.

A crystal structure of NahE was solved to 1.9 Å resolution with active site electron density corresponding to BP and the products benzaldehyde and pyruvate. The structure was obtained by soaking NahE crystals in BP. When the structure was modeled to include BP (**NahE-BP**, **Figure 2-17A**), C-4 is much further from the hydroxyl group of Tyr-155 than the distance in NahE-HN (**Figure 2-25**). The $2F_o$-F_c electron density maps show decreased density between C-3 and C-4 of BP in NahE-BP, indicating that NahE has processed BP. The strcture was modeled to indlude the pyruvate-Lys-183

Schiff base moiety and benzaldehyde in a separate refinement (**NahE-BPy**). When the pyuvoyl adduct and benzaldehyde were included in modeling, the carbonyl oxygen of benzaldehyde is hydrogen bonded to the side chain oxygen atoms of Asn-281 and Glu-285.

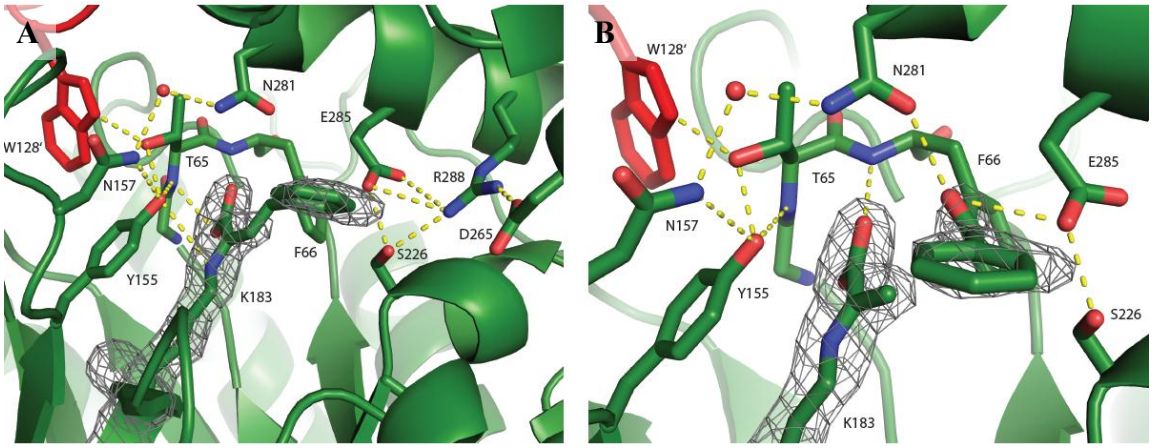


Figure 2-17 The 1.9 Å resolution crystal structures of NahE-BP and NahE-BPy

The same experimental data was used for modeling of **A)** BP adduct (NahE-BP) and **B)** pyrovoyl adduct and product benzaldehyde (NahE-BPy). $2F_o$-F_c electron density surrounds the ligands, contoured at 1.0 σ.

Data Collection	NahE-Sal	NahE-BP	NahE-BPy
space group	$P3_221$	$P3_221$	
cell dimensions			
a, b, c (Å)	88.136, 88.136, 141.19	84.609, 84.609, 132.53	
α, β, γ (deg)	90, 90, 120	90, 90, 120	
resolution (Å)	2.88-40.00	1.90-50.00	
R_{sym} (%) (linear R factor)	0.116 (0.435)	0.084 (0.549)	
I/σ	11 (2.9)	18 (3.1)	
completeness (%)	97.4 (96.5)	100.0 (100.0)	

Refinement	NahE-SAL	NahE-BP	NahE-BPy
resolution (Å)	40.06-2.91	49.15-1.90	49.15-1.90
no. of reflections	14261	43849	43849
R_{work}/R_{free} (%)	19.31/24.96	16.35/19.98	15.75/19.35
no. of atoms			
protein	5020	5044	5030
water	22	358	383
Average *B* factor (Å^2)			
protein	46.0	36.7	36.4
water	29.0	39.8	40.5
rmsd			
bond lengths (Å)	0.003	0.006	0.010
bond angles (deg.)	0.708	0.960	1.130
Ramachandran plot (%)			
favored regions	96.3	98.0	98.0
allowed regions	3.7	2.0	2.0
disallowed regions	0.0	0.0	0.0

Data for the last resolution shell are given in parentheses.

Table 2-10 Crystallographic data and refinement statistics for NahE-Sal, NahE-BP, and NahE-BPy.

2.3.8 Crystal Structures of PhdJ

The crystal structure of PhdJ was solved in order to determine a structural basis for the substrate preference (i.e. CBP vs. HBP) and to better understand its ability to activate water. Structures of native PhdJ and PhdJ in complex with CBP, the 4-hydroxy intermediate, CBA, and pyruvate were obtained by soaking experiments. Phases for all PhdJ structures were solved using molecular replacement. NahE was used as a search model for PhdJ, and all PhdJ complexes were solved using native PhdJ as a search model. All PhdJ crystals grew in the $P2_1$ space group with four monomers per asymmetric unit. Like NahE, the enzyme is a tetramer and shows the primary active site residues conserved in the NAL subgroup: Lys-180, Tyr-152, and the GXXGE motif, where the XX residues are Thr-65 and Phe-66 (**Figure 2-18A,B,D**). Trp-125′, Asn-154, Ser-278, and Asp-282 are secondary residues, although Asn-154 is conserved in NahE. Although the monomeric structure deviates slightly from DHDPS, the triose phosphate isomerase barrel and C-terminal helices remain intact (RMSD for PhdJ and DHDPS is 4.3 Å and the RMSD for PhdJ and NAL is 1.7 Å, **Figure 2-18C**).

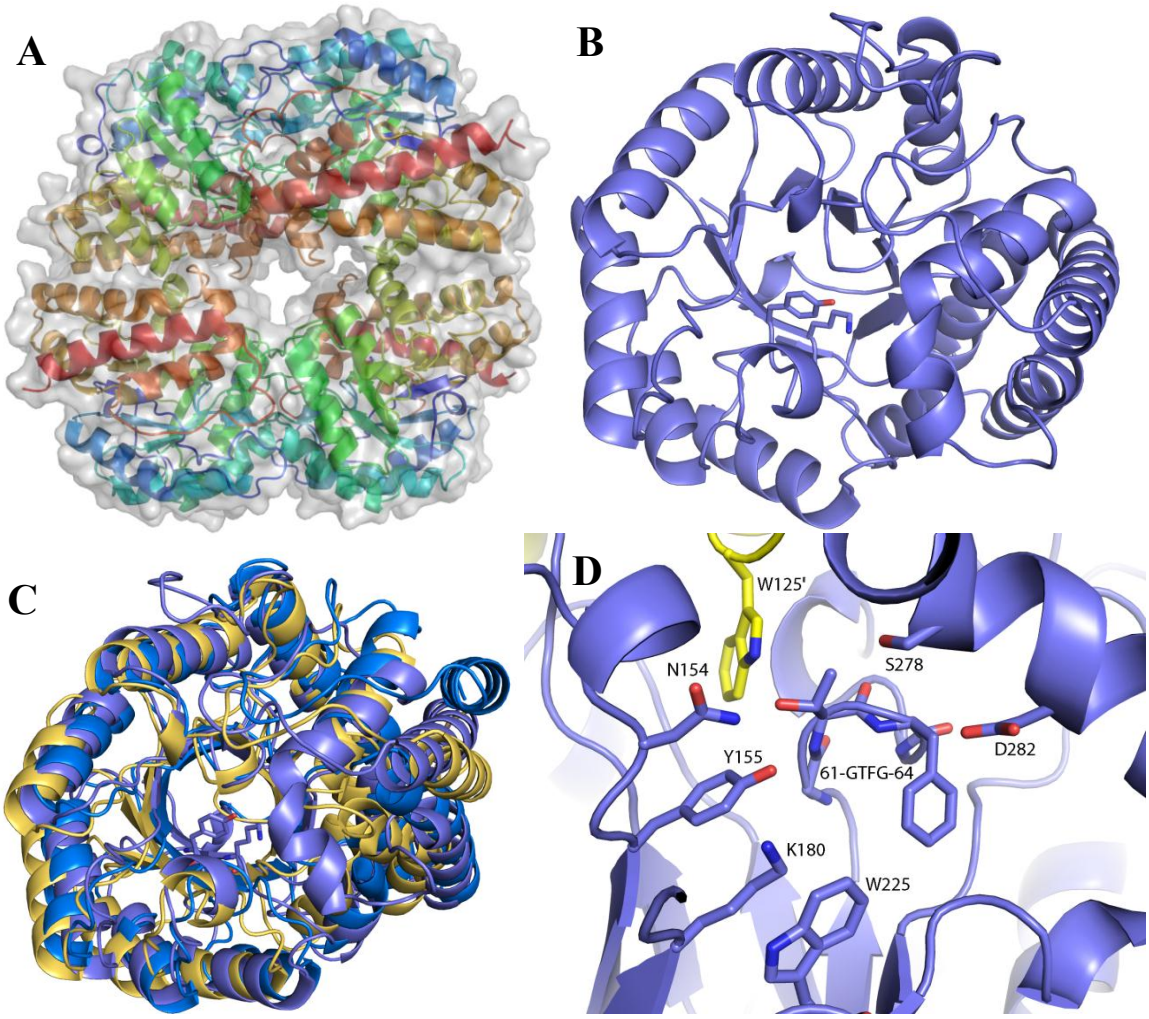
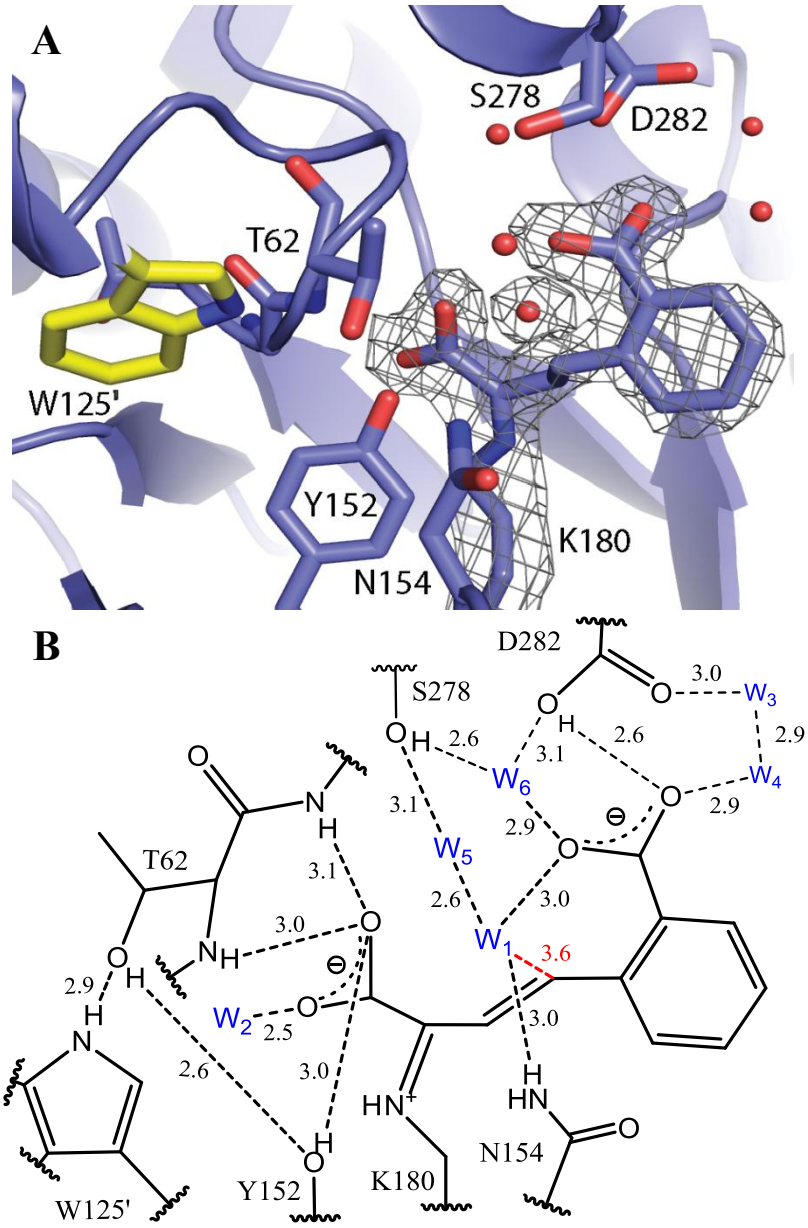


Figure 2-18 The 2.1 Å resolution crystal structure of PhdJ.

A) The PhdJ tetramer (rainbow gradient from blue N-terminus to red C-terminus) with transparent surface shown. **B)** PhdJ monomer with Lys-180 and Tyr-152 shown in sticks. **C)** Alignment of PhdJ (slate), NAL (marine, PDB entry 1NAL), and DHDPS (yellow orange, PDB entry 1DHP) **D)** Active site of PhdJ.

The structure of PhdJ in complex with CBP (PhdJ-CBP) was obtained by soaking experiments. The 2.0 Å structure has 4 monomers per asymmetric unit where two active sites show CBP and the two other active sites have pyruvate and CBA. The interaction of the C-1 carboxylate of CBP with active site residues of PhdJ is similar to those observed for NahE, DHDPS, NAL, and KDGA. A C-1 carboxylate oxygen hydrogen bonds with the backbone amide hydrogens of Thr-62 and Phe-63, as well as the side chain hydroxyl of Tyr-152 (**Figure 2-19A,B**).

A water molecule (W_1) is 3.6 Å from the benzylic carbon, near the C-4 carbon where addition occurs. Its position is, however, in plane with the double bond, implying that a conformational change must occur before hydration, so that the water is added on the *Re*-face, as deduced by the configuration of the 4-hydroxy intermediate (**Figure 4-20**). W_1 is coordinated by W_5, the side chain nitrogen of Asn-154, and the *o*-carboxylate oxygen of the substrate. If W_1 is adds to C-4 of the CBP adduct, both W_5 and the carboxylate group are capable of accepting a proton from W_1. Protonation at C3 might be facilitated by the side chain of Tyr-152 (4.1 Å from C-4) or the *o*-carboxylate (3.4 Å from C-4) (**Section 2.4.3**). A small helix at the C-terminus of the β-barrel resembles the helix observed in the NahE structures, and might seclude the active site from solvent. The helix has four nonpolar residues (Tyr-266, Phe-271, Phe-274, and Leu-275) on the face of the helix near Lys-183, which might increase the hydrophobicity of the binding pocket (**Figure 2-19C**).


A
S278
D282
T62
W125'
Y152
K180
N154
B
D282
S278
T62
W125'
Y152
K180
N154
W1
W2
W3
W4
W5
W6
3.0
2.6
3.1
2.9
3.6
2.5

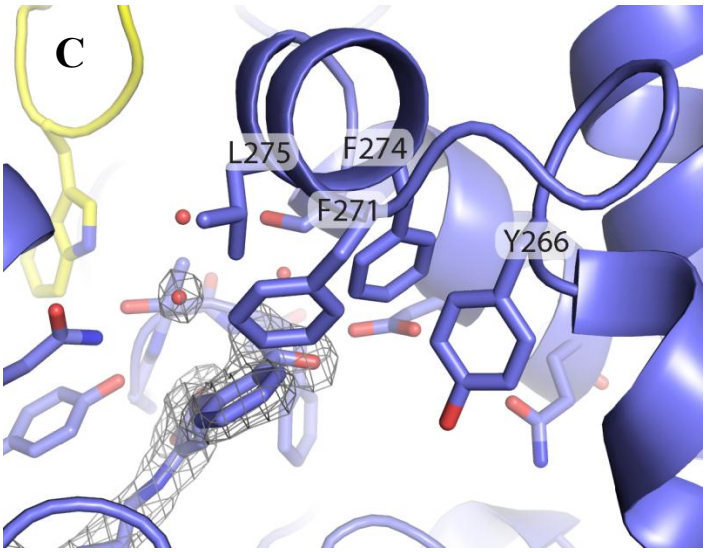


Figure 2-19 Active site of the 2.0 Å resolution PhdJ-CBP crystal structure**.**

A) Continuous $2F_o - F_c$ electron density, contoured at 1.0 σ, surrounding Lys-183 and CBP. **B)** Diagram of interactions near CBP with hydrogen bonds represented as dashes, except for the red dash from W_1 the the benzylic carbon of CBP. Distances are in Å. **C)** Hydrophobic residues of α-helix 10 in the active site.

The structure of PhdJ in complex with (*R*)-2-(3-carboxy-1-hydroxy-3-oxopropyl)benzoate (CN, for *o*-carboxy intermediate) was solved to 2.0 Å resolution via after soaking crystals with CBP (PhdJ-CN, **Figure 2-20B**). The hydroxyl oxygen of the intermediate is hydrogen bonded to the side chain nitrogen of Asn-154 and side chain hydroxyl of Tyr-152 (**Figure 2-20A,B**). The *o*-carboxyl group of CN is positioned perpendicular the aromatic ring, and closer to the side chain of Asp-282 than is observed for the CBP adduct in PhdJ-CBP.

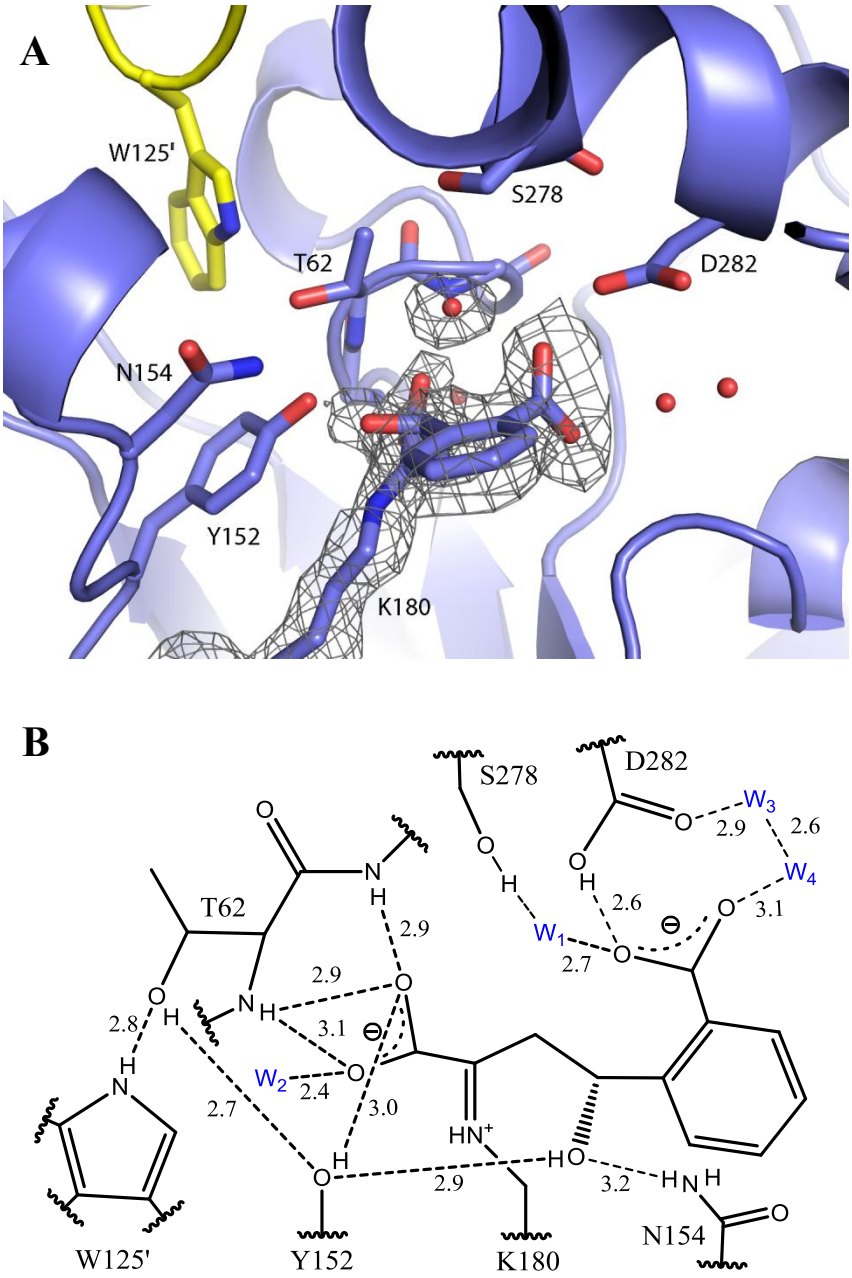


Figure 2-20 Active site of the 2.0 Å resolution PhdJ-CN crystal structure.

A) $2F_o$-F_c electron density map (0.5 σ) surrounding CN, Schiff base linkage with Lys-180, and W_1. **B)** Diagram of interactions with CN. Distances are in Å.

Two monomers of PhdJ-CBP contain CBA and pyruvate (**Figure 2-21A**). The interactions of the *o*-carboxylate oxygen atoms of CBA with the side chain of Asp-282 are similar to that observed in PhdJ-CN (**Figure 2-21B**).

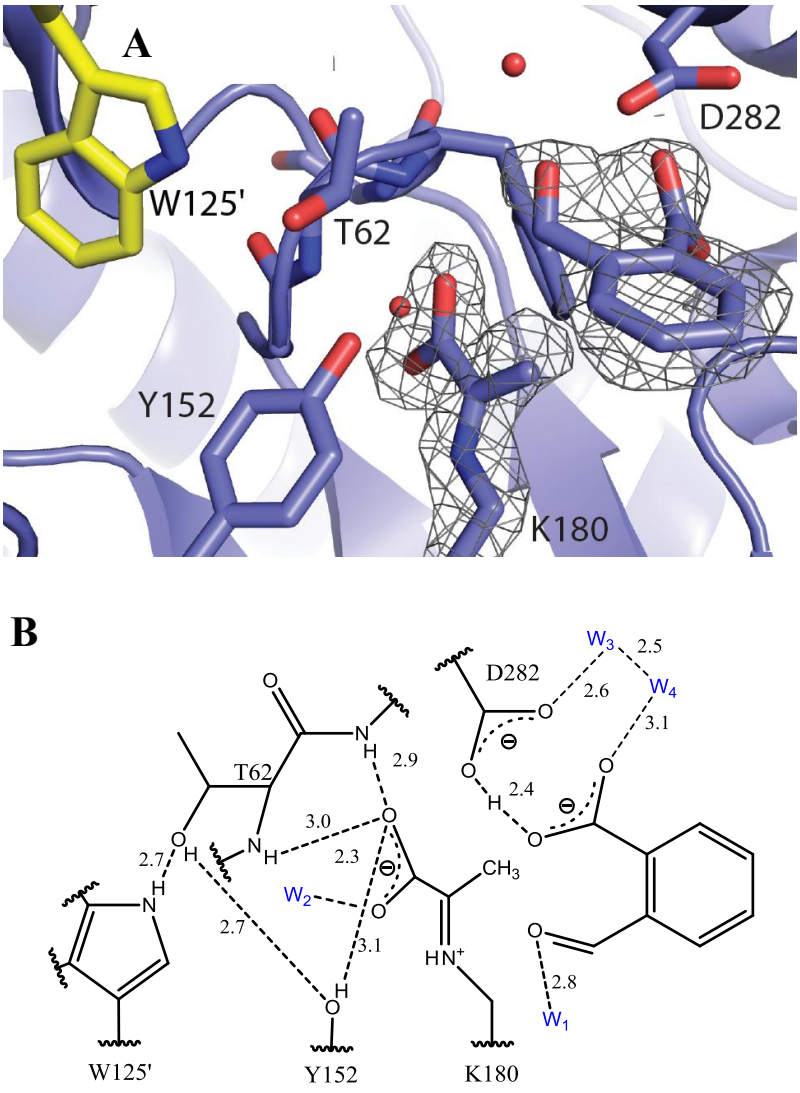


Figure 2-21 Active site of PhdJ-CBP with pyruvate and CBA.

A) $2F_o$-F_c continuous electron density surrounding Lys-183, the pyruvoyl adduct, and CBA. **B)** Diagram of interactions near products in active site with distances in Å.

A crystal structure PhdJ attached to the benzylic carbon of CBA (**PhdJ-CBA**) was determined after soaking experiments using CBA. The 2.2 Å resolution crystal structure shows that Lys-180 forms Schiff base with CBA in the active site. The carboxylate group of the ligand is pointed outwards toward the active site opening and is coordinated to side chain amide nitrogen of Asn-154 and a water molecule. The hydrophobic ring is sandwiched between Thr-62 and Phe-63 in the GXXGE motif.

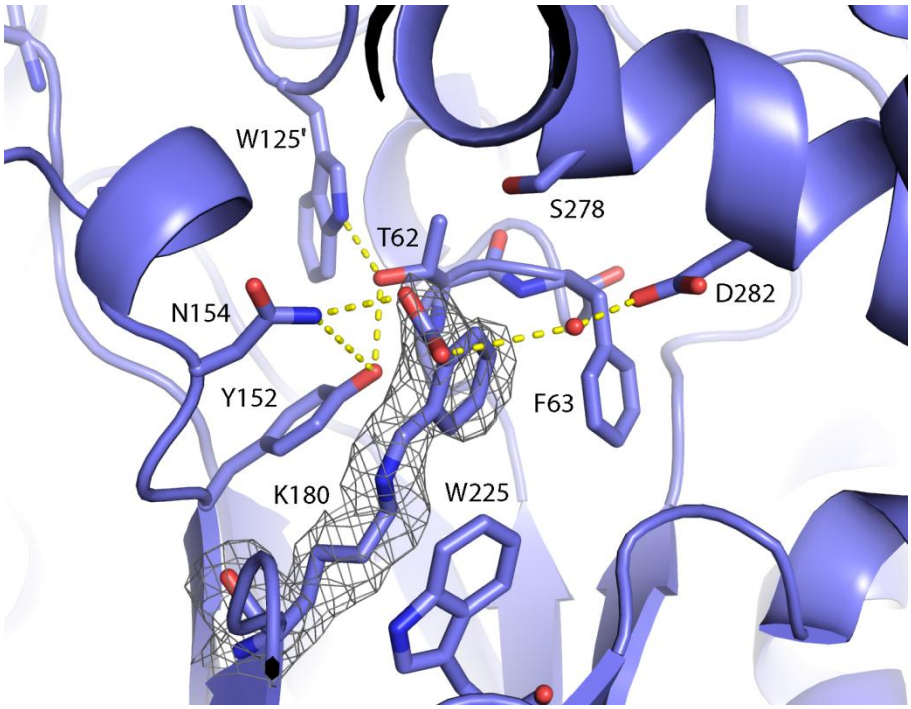


Figure 2-22 Active site of the 2.2 Å resolution PhdJ-CBA crystal structure with $2F_o$-F_c electron density map (0.5 σ) of surrounding CBA and the Schiff base linkage to Lys-180.

Data Collection	PhdJ	PhdJ-CBP	PhdJ-CN	PhdJ-CBA
space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
cell dimensions				
a, b, c (Å)	52.4, 137.0, 88.3	52.5, 137.2, 88.3	52.5, 137.2, 88.4	52.3, 136.9, 88.2
α, β, γ (deg)	90.0, 99.1, 90.0	90.0, 98.9, 90.0	90.0, 98.9, 90.0	90.0, 99.0, 90.0
resolution (Å)	40.00-2.05	50.00-2.00	40.00-2.00	2.06-50.00
R_{sym} (%) (linear R factor)	0.104 (0.496)	0.087 (0.326)	0.082 (0.222)	0.116 (0.492)
$I/\sigma I$	11 (2.1)	14 (3.2)	14 (4.5)	11 (2.5)
completeness (%)	94.3 (83.6)	93.5 (88.7)	99.9 (99.9)	100.0 (99.9)
Refinement				
resolution (Å)	39.90 - 2.05	45.30-2.00	39.32-2.00	39.85-2.21
no. of unique reflections	72626	77202	82301	61343
R_{work}/R_{free} (%)	17.95/22.39	15.21/18.93	17.03/19.96	16.84/20.58
no. of atoms				
protein	9893	9960	9997	9985
water	894	1116	1145	778
Average *B* factor (Å^2)				
protein	32.9	21.8	19.1	23.8
water	39.1	32.9	31.9	31.2
rmsd				
bond lengths (Å)	0.002	0.007	0.0023	0.004
bond angles (deg)	0.644	1.023	0.6923	0.814
Ramachandran plot (%)				
favored regions	98.15	98.69	98.46	98.00
allowed regions	1.85	1.31	1.54	2.00
disallowed regions	0.00	0.00	0.00	0.00

[a]Data for the last resolution shell are given in parentheses.

Table 2-11 Crystallographic data and refinement statistics for PhdJ, PhdJ-CBP, PhdJ-CN, and PhdJ-CBA.

2.3.9 Crystal Structures of PhdG

The crystal structure of PhdG was determined to assist the identification of its biological function, as well as to determine the structural basis for its substrate preference and mechanism. The identification for HBP as a substrate (April 2017) as a substrate via UV/Vis spectrophotometry sheds light on structure function relationships when comparing hydratase-aldolases. Crystals grew in the $P4_32_12$ space group with two monomers in the asymmetric unit. The 2.7 Å resolution PhdG crystal structure was solved by molecular replacement using an unpublished native NahE structure as a search model. Monomeric backbone deviation after superpositioning of PhdG to NAL and DHDPS was low (RMSD values of 1.82 and 1.44, respectively, **Figure 2-23A,B**).

Lys-180, Tyr-152, and the GXXGE motif are conserved in PhdG, like all other NAL subgroup members. The XX residues in the GXXGE motif are Ser-62 and Ala-63, which differ from Thr and Phe residues found in NahE and PhdJ (**Figure 2-23C**). The side chain hydroxyl of Tyr-125 makes the intersubunit interaction with the side chain of Ser-62, like the equivalent tyrosine residues do in NAL, DHDPS, and KDGA, but unlike NahE, PhdJ (both use Trp), and KDGDH (which doesn't make an equivalent hydrogen bond). Asn-154 and Trp-221 in PhdG are conserved in NahE and PhdJ, but are not conserved in other NAL subgroup members. Thr-182 is not conserved with other NAL subgroup enzymes and the side chain oxygen hydrogen bonds with the side chains of Lys-180. Gln-269 and Gln-272 are also nonconserved residues that are proximal to Lys-180.

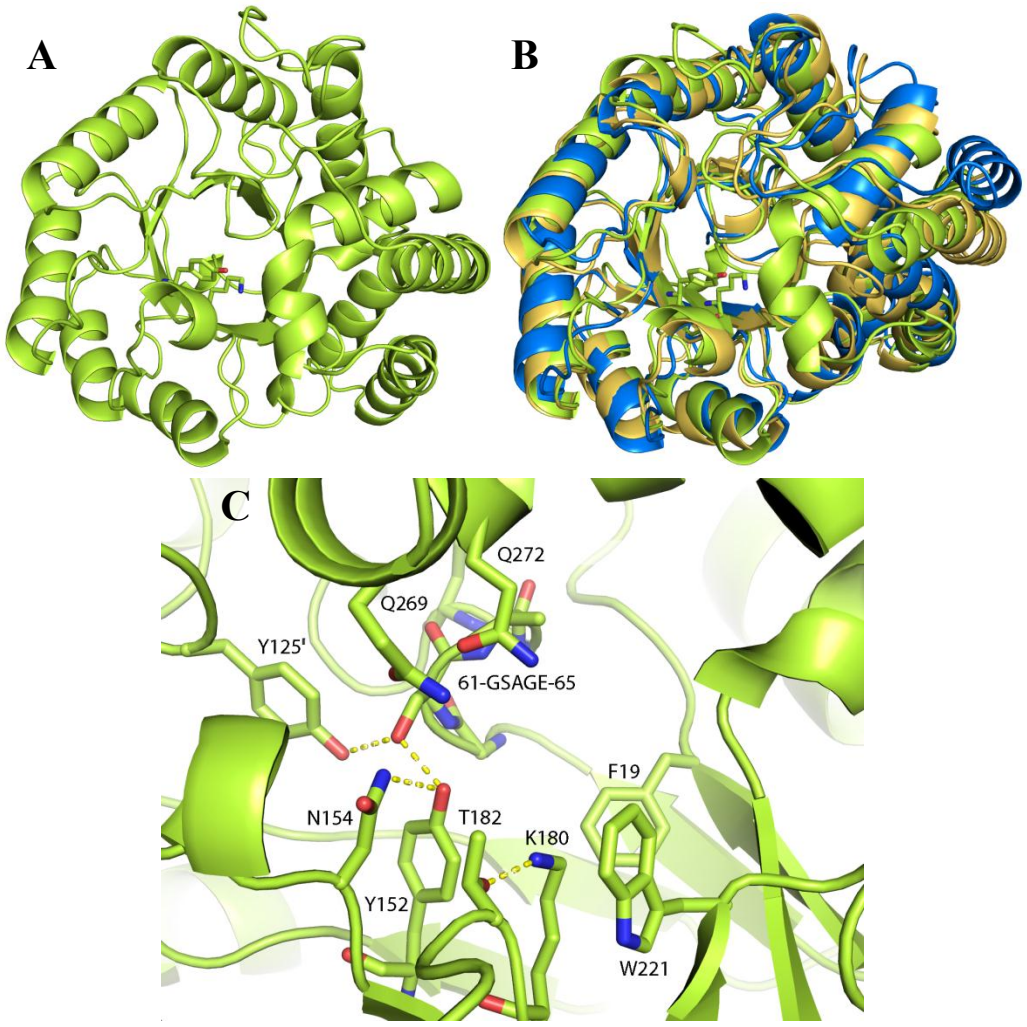


Figure 2-23 The 2.7 Å resolution crystal structure of PhdG.

A) PhdG monomer. **B)** Alignment of PhdG (limon), NAL (marine, PDB entry 1NAL), and DHDPS (yellow orange, PDB entry 1DHP). **C)** Active site of PhdG.

A 1.9 Å resolution crystal structure of PhdG in complex with a pyruvate carbinolamine (PhdG-Pyr) was determined after soaking experiments using pyruvate. The hydroxyl of the carbinolamine hydrogen bonds with the hydroxyl group of Tyr-152

(**Figure 2-24A**) and is an intermediate in Schiff base formation. A water molecule bridges the side chains of Gln-269 (Nε) and Gln-272 (Oε) and contacts the carbinolamine hydroxyl group. PhdG has numerous nonpolar residues nearby including Leu-205, Trp-221, Ile-223, Phe-258, Met-265, Phe-268, Trp-275, and Trp-276 (**Figure 2-24B**).

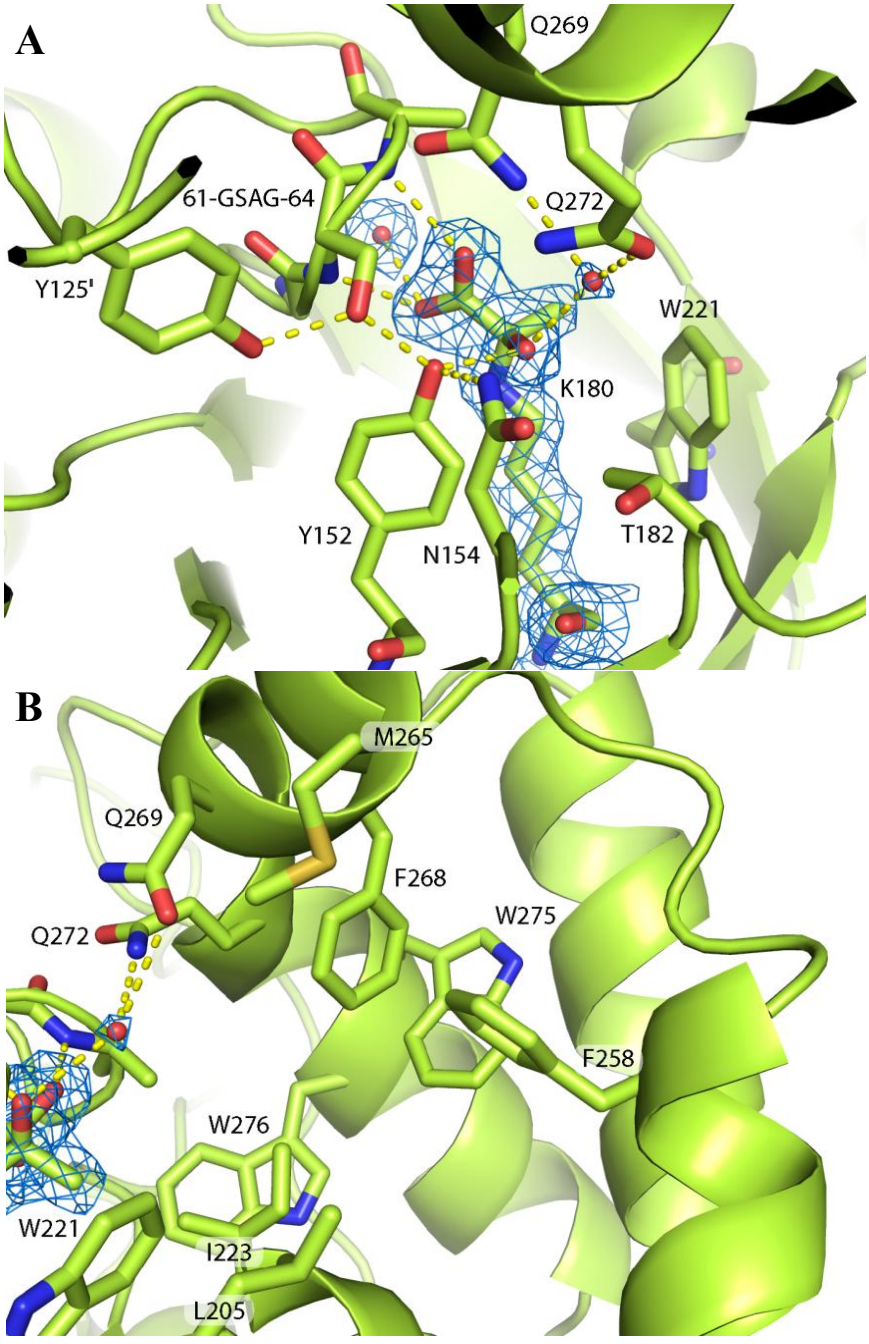


Figure 2-24 Active site of the 1.9 Å resolution PhdG-Pyr crystal structure.

A) $2F_o$-F_c map of continuous density surrounding Lys-180 and carbinolamine adduct, countoured at 1.0 σ. **B)** Hydrophobic residues near the active site of PhdG.

Data Collection	PhdG	PhdG-Pyr
space group	$P4_32_12$	$P4_32_12$
cell dimensions		
a, b, c (Å)	86.879, 86.879, 183.907	86.671, 86.671, 184.133
α, β, γ (deg)	90.000, 90.000, 90.000	90.000, 90.000, 90.000
resolution (Å)	50.00 - 2.65	50.00 - 1.94
R_{sym} (%) (linear R factor)	1.041 (0.247)	1.105 (0.168)
$I/\sigma I$	13 (3.0)	17.6 (2.1)
completeness (%)	100.0 (100.0)	100.0 (100.0)
Refinement		
resolution (Å)	43.440 - 2.650	43.3461 - 1.9378
no. of reflections	21207	52924
R_{work}/R_{free} (%)	23.42/28.12	16.80/19.53
no. of atoms		
protein	4904	4908
water	159	789
Average *B* factor (Å^2)		
protein	33.50	21.60
water	26.10	29.05
rmsd		
bond lengths (Å)	0.003	0.005
bond angles (deg)	0.738	0.983
Ramachandran plot (%)		
favored regions	98.43	98.58
allowed regions	1.57	1.42
disallowed regions	0.00	0.00

Data for the last resolution shell are given in parentheses.

Table 2-12 Crystallographic data and refinement statistics for PhdG and PhdG-Pyr.

2.4 DISCUSSION

2.4.1 Hydratase-Aldolases in the NAL Subgroup

The hydratase-aldolases NahE, PhdJ, and PhdG were identified as members of the NAL subgroup of the Type I aldolase superfamily by sequence analysis. As such, they have structural traits from the NAL subgroup including the $(\alpha/\beta)_8$ barrel, C-terminal helices, the conserved Lys-183 and Tyr-155 residues (NahE numbering), and the GXXGE motif (LeVieux, *et al.* 2016). The crystal structures are highly homologus, as indicated by their Cα RMSD values, which range from 0.72 to 0.94 Å. DLS experiments show that all three enzymes are tetrameric, like other members of the NAL subgroup. ESI-MS, chemical mutagenesis, and crystallographic analysis indicate the enzymes form Schiff base adducts with their α-keto acid substrates through the conserved lysine. Crystal structures of each enzyme with bound ligand show the C-1 carboxylate of the substrates interacts with the backbone amide hydrogens of the XX residues in the GXXGE motif.

NahE, PhdJ, and PhdG have several other residues that are not shared with other NAL subgroup members, which are referred to as secondary residues. These residues are likely responsible for the distinct chemistry of the hydratase-aldolases. The enzymes have Asn-157 (NahE numbering) separated by two amino acids from Tyr-155 (sequence alignment, **Figure 2-25**). In NahE and PhdJ, this residue is near the benzylic carbon of benzylidenepyruvate substrates bound in crystal structures, to which water is added. The asparagine residue interacts with a water molecule in both the NahE-HN and PhdJ-CBP structures (**Figure 2-14, 2-19**). In NahE-HN, the side chain amide nitrogen Asn-157 is

hydrogen bonds with a water molecule that interacts with the amide nitrogen of Asn-281 above the 4-hydroxyl intermediate HN (**Figure 2-14**). In PhdJ-CBP, the side chain amide nitrogen of Asn-154 hydrogen bonds to a water molecule only 3.6 Å from the benzylic carbon where water is added. The same water molecule is in contact with one *o*-carboxylate oxygen atom of CBP and another water molecule (**Figure 2-19**). The Asn might be involved in positioning the water molecule and stabilization of the transient hydroxide ion during addition in NahE and PhdJ, and sets precedence for the role of the conserved asparagine in PhdG and other hydratase-aldolases.

Trp-224 (NahE numbering) is another secondary residue located on β-strand 8. The side chain is parallel to the side chain of Lys-183 in crystal structure of NahE, PhdJ, and PhdG. The residue may lower the pK_a of the amine group of Lys-183 to increase nucleophilicity for Schiff base formation, or it may contribute to positioning of the substrate. Other important amino acids are specific to the individual hydratase-aldolases and are referred to as tertiary residues (**Figure 2-25**).

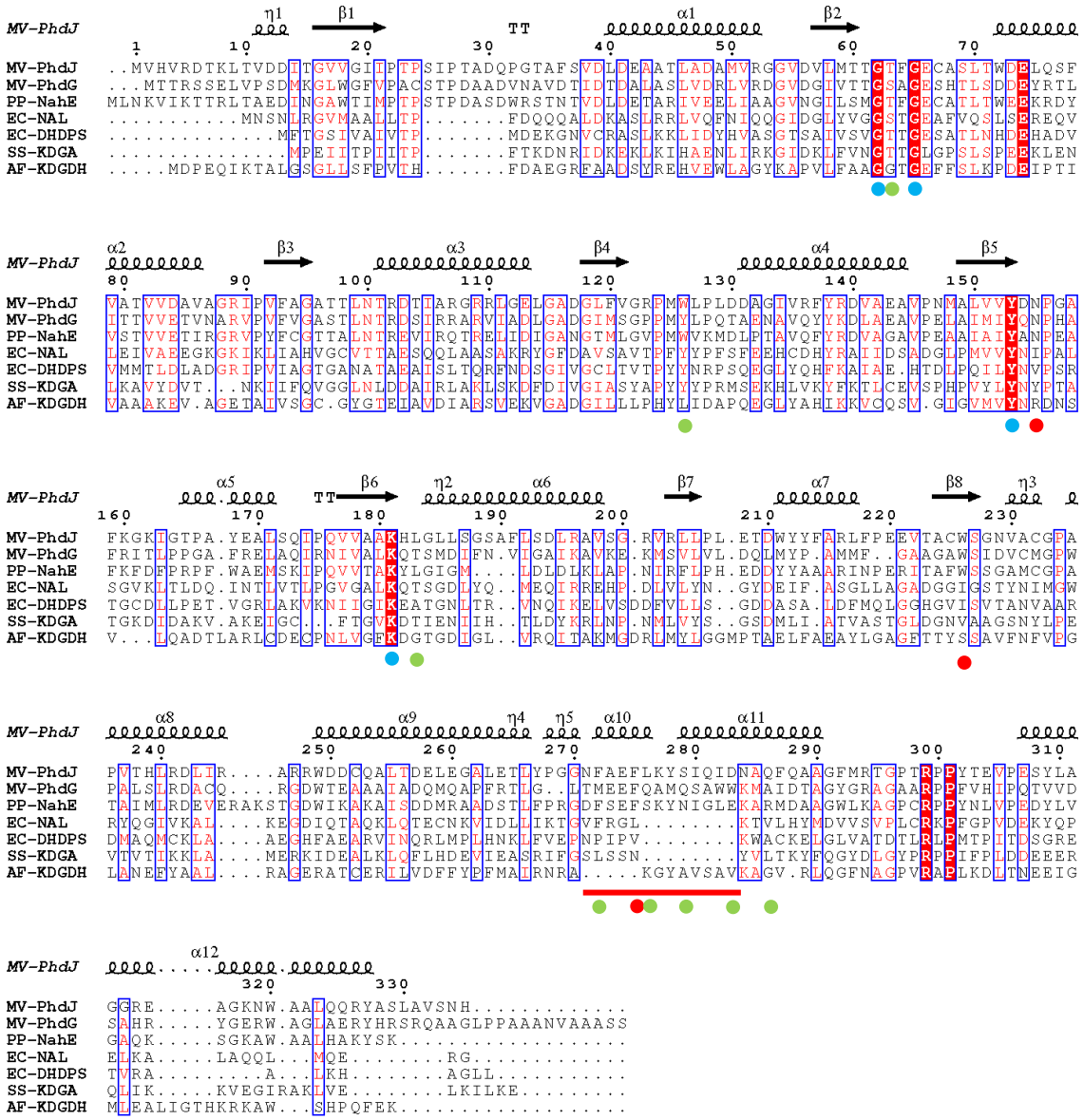


Figure 2-25 Sequence alignment of representative NAL subgroup members.

Secondary structure for PhdJ is shown above the alignment. The dots beneath select residues indicate primary residues (blue) in the NAL subgroup, secondary residues (red) common in hydratase-aldolases, and tertiary residues (green) that are specific to individual hydratase-aldolases. The red line indicates an insertion in hydratase-aldolases that which codes for α10.

When sequences of hydratase-aldolases are aligned and compared to other members of the NAL subgroup, the hydratase-aldolases contain an insertion near position 270 (**Figure 2-25**). The 8-residue insertion in NahE, PhdJ, and PhdG allows for an α-helix (designated α10) to form between two longer α-helices at the C-terminus. This α-helix extends over the C-terminal opening of the β-barrel forming a deep cavity, a feature not seen in other NAL subgroup members (**Figure 2-26A**). Beneath the helical "shield" are hydrophobic side chains that could assist in the binding of the less polar moieties of substrates for these 3 enzymes (**Figure 2-26B**). A nonpolar environment in the active site would explain the high affinity of hydratase-aldolases for their substrates, as the K_M values are in the mid to low μM range, whereas those for DHDPS and NAL are in the high μM to low mM range range (Aisaka *et al.*, 1991, Laber *et al.*, 1992). One hydrophobic residue Phe-277 is located on α10 and is conserved in NahE, PhdJ, and PhdG, and can be considered a secondary residue (**Figure 2-26B**). Other key residues identified in the crystal structures reported here vary among NahE, PhdJ, and PhdG (tertiary residues) and are discussed in the context of each enzyme in the following sections.

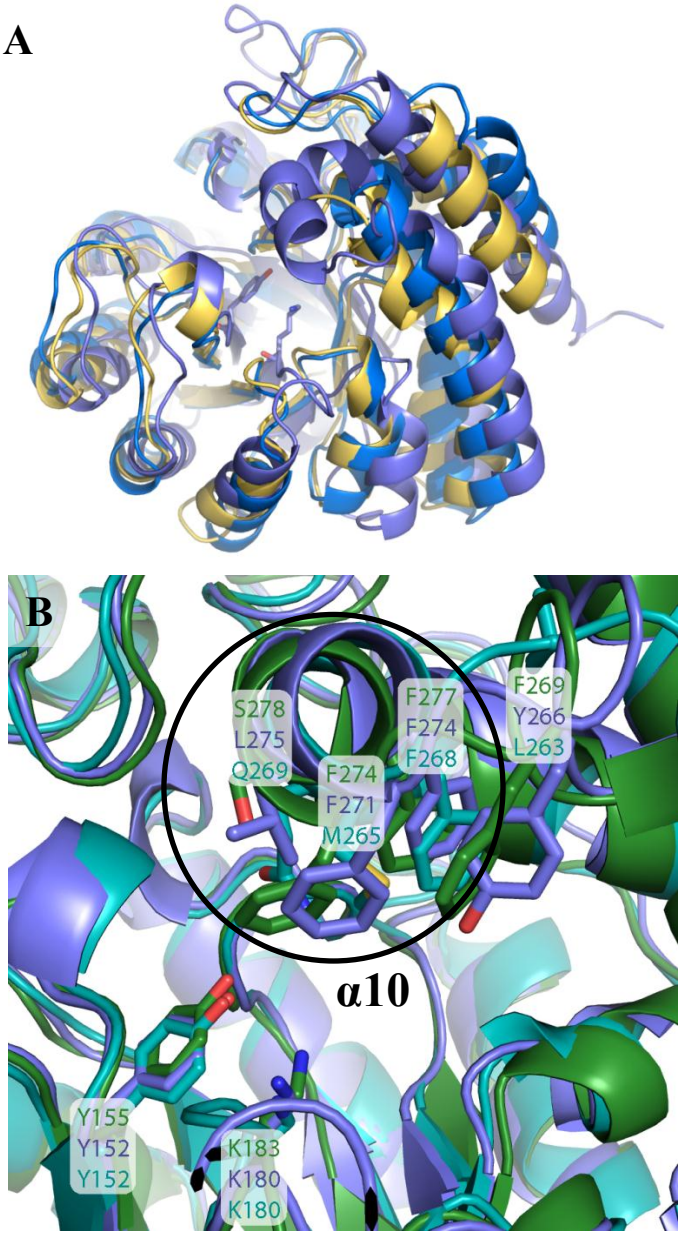


Figure 2-26 The helix α10 unique to hydratase-aldolases in the NAL subgroup.

A) Alignment of PhdJ (slate), NAL (marine, PDB entry 1NAL), and DHDPS (yellow orange, PDB entry 1DHP). **B)** Close-up of α10 in NahE (forest), PhdJ (slate), and PhdG (light teal). Residues on the helix are shown in reference to the conserved lysine and tyrosine residues.

2.4.2 The Mechanism of NahE Using HBP

NahE exhibits a strong preference for HBP, with the k_{cat}/K_M value (1×10^7 $M^{-1}s^{-1}$) 10^4-fold greater than that using BP or CBP, in part due to its low K_M (1.2 μM), which is ~100-fold less than the value measured for BP or CBP. Other reports have shown relative rates with BP and CBP that are close to 1% of that using HBP (Kuhm *et al.*, 1993, Eaton, 2000). In addition, the reported K_M values are in the low μM concentration range (Kuhm *et al.*, 1993), which is consistent with our results.

The k_{cat} values for NahE catalysis in the reverse direction using salicylaldehyde and pyruvate (18.7 s^{-1}) were similar to that using HBP (11.7 s^{-1}). The K_M value for pyruvate was high (6 mM), suggesting low affinity of NahE and pyruvate, despite the favorable interactions observed in the crystal structure in complex with pyruvate (Schiff base) and salicylaldehyde (NahE-SP, **Figure 2-15**). This structure shows Schiff base formation between C-2 of pyruvate and the side chain nitrogen of Lys-183 and hydrogen bonds between the C-1 carboxylate of pyruvate and backbone of the GXXGE motif. The K_M value with respect to pyruvate is significantly lower for *E. coli* DHDPS (0.26 mM) (Dobson *et al.*, 2004) and PhdG (0.27 mM).

Crystal structures were obtained for native NahE and NahE in complex with the 4-hydroxy intermediate, saicylaldehyde and pyruvate, salicyaldehyde (Schiff base), and BP (or products benzaldehyde and pyruvate). The overall crystal structure of NahE did not change significantly when in complex with HN, BP, or salicylaldehyde and pyruvate (Cα RMSD < 0.3 Å, **Figure 2-25A**). In addition to overall homology of tertiary structure,

the active site residues remained in almost identical positions in the unliganded and liganded structures (**Figure 2-25B,C**).

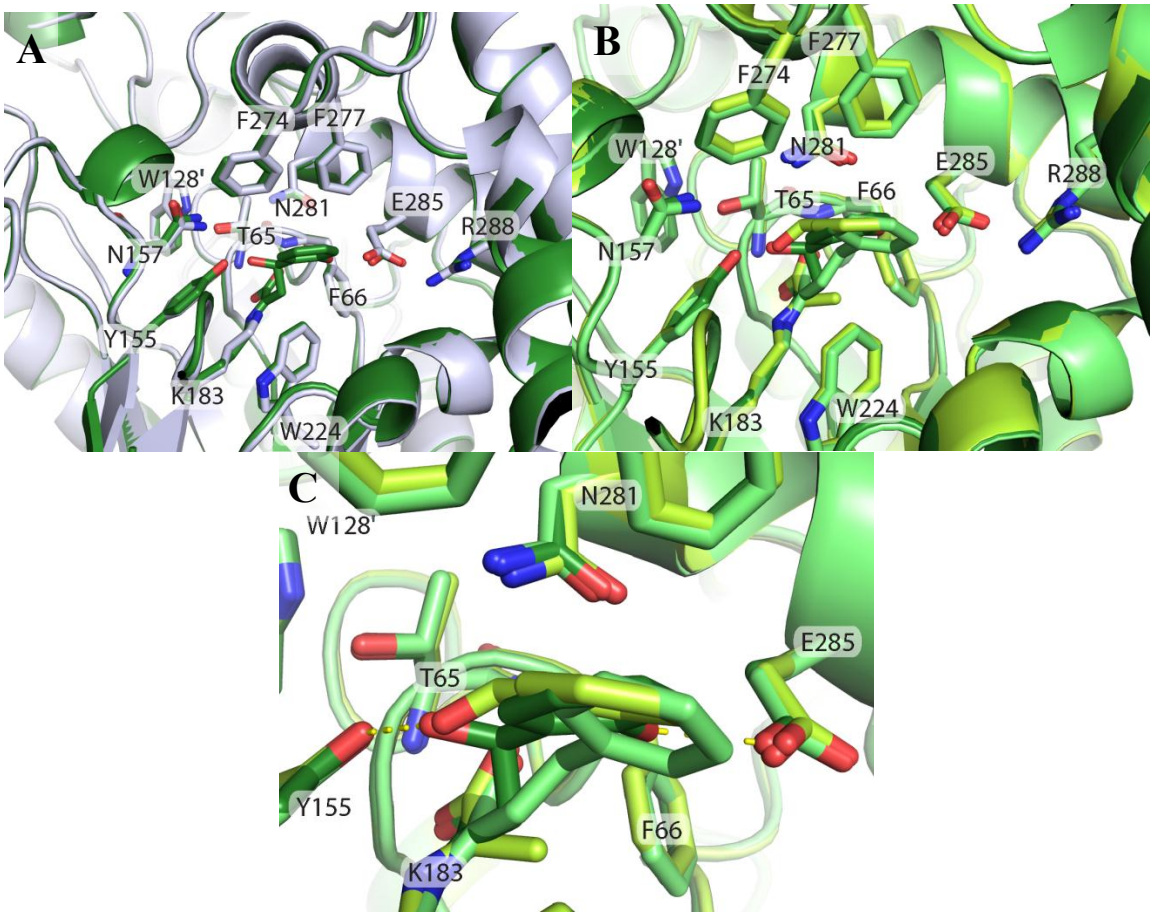


Figure 2-27 Superposition of NahE crystal structures.

A) NahE (grey) alignment with NahE-HN (forest). **B)** Superposition of NahE-HN (forest), NahE-BP (lime), and NahE-SP (limon). **C)** Close-up of ligands shown in **B**.

Based on the crystal structures of NahE and mechanisms in the NAL subgroup, the following catalytic mechanism is proposed (**Scheme 2-2**). First, Schiff base formation between the side chain nitrogen of Lys-183 and C-2 of HBP occurs, mediated by the

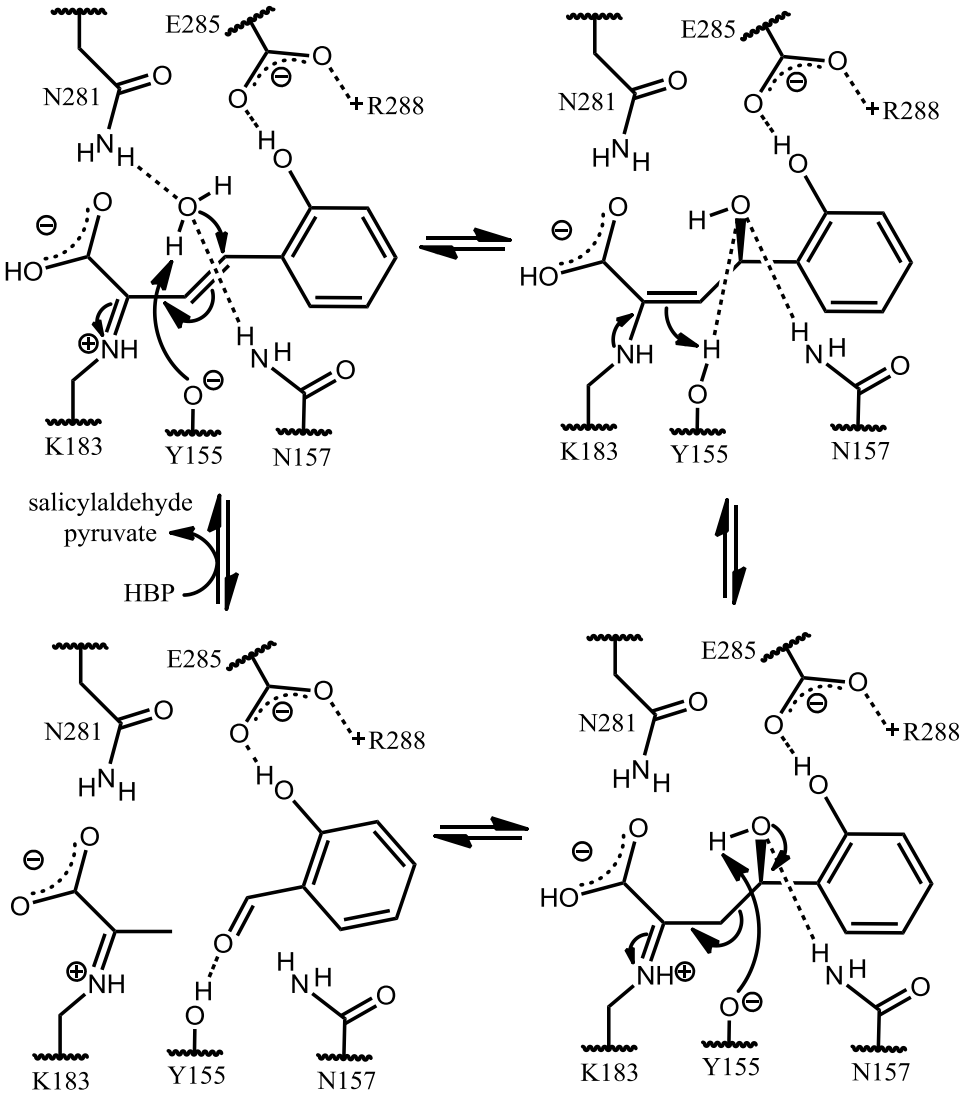
hydroxyl group of Tyr-155, as suggested for other type I aldolases (Choi *et al.*, 2006), based on structure analysis and kinetic studies. Crystal structures of fructose-1,6-bisphosphate aldolase (Lorentzen *et al.*, 2005) and PhdG (**Section 2.4.5**) have covalently bound carbinolamine species (intermediates in Schiff base formation and decay) where the C-2 oxygen hydrogen bonds to the side chain hydroxyl group of Tyr-152 (PhdG numbering), suggesting the residue mediates proton transfer and hydration/dehydration during reactions in these enzymes. Subsequent to Schiff base formation, the side chain amide nitrogen of Asn-157 (and possibly the same nitrogen of Asn-281) position the water molecule, and Tyr-155 acts as a general base for the addition of water to the *Re*-face of HBP. Evidence for the step comes from the NahE-HN structure which shows the C-4 oxygen of the intermediate HN hydrogen bonds with the side chain amide nitrogen of Asn-157 and side chain oxygen of Tyr-155 (**Figure 2-14**). It is possible the side chain amide nitrogen Asn-281 is also involved, though the distance to the 4-hydroxyl group is 3.8 Å, whereas the distance from Asn-157 is 3.3 Å. Asn-278 is the counterpart to Asn-281 conserved in DHDPS (sequence alignment in **Figure 2-25**, Asn-281 on left side of gap). Electrons are pushed to the side chain nitrogen of Lys-183, forming an enamine, prior to protonation at C-3 by the side chain oxygen of Tyr-155 (*syn*- addition). A single general acid/base for *syn*- addition is proposed for Enoyl-CoA hydratase (Bahnson *et al.*, 2002), where the proposed E1cb mechanism features a carbanion/enolate intermediate. The reason a quinone methide mechanism is not operative (as suggested by rate comparison using HBP, and its *o*- and *m*- derivatives discussed in **Section 1.2.3**, Eaton, 2000) may be due to the hydrogen bond donation by the *o*-hydroxyl group of HBP to the

carboxylate oxygen of Glu-285. Glu-285 is likely deprotonated as one carboxylate oxygen of Glu-285 interacts with the side chain of Arg-288 (Nη), probably in a charge-charge interaction. Because the *o*-hydroxyl group is not deprotated to form a quinone methide, the functional group draws electron density by induction from the benzylic carbon, making the atom more electrophilic and susceptible to Micheal addition.

Deprotonation during water addition by the *ortho* phenoxide of HBP cannot be ruled out (2.8 Å in NahE-HN), but a dramatic change in conformation would need to occur after addition at the *Re* face of HBP, to produce the (*R*)-4-hydroxyl adduct observed in NahE-HN. Additionally, the interaction of Glu-285 and Arg-288, and the evidence against a quinone methide intermediate suggest the *ortho*-hydroxyl group of HBP is not deprotonated during the reaction.

Another reasonable scenario for water activation includes the C-1 carboxylate group of HBP. One oxygen of this group is only 3.1 Å (in NahE-HN) from the benzylic position of the intermediate HN, and also interacts with the side chain hydroxyl of Tyr-155 (3.2 Å, NahE-HN). The distance is close enough to add water, but the C-1 carboxlate is in plane with the C-3/C-4 double bond of BP in (NahE-BP, **Figure 4-17**), where Tyr-155 is on the *Re* face of the C-4, to which water is added. Water activation by carboxylate groups (glutamate side chains) for hydration of α,β-unsaturated substrates has been reported in Enoyl-CoA hydratase, hydroxycinnamoyl-CoA hydratase-lyase (Bahnson *et al.*, 2002, Bennett *et al.*, 2008), and tautomerase superfamily members such as *cis*-CaaD and MSAD (Poelarends *et al.*, 2003, Huddleston *et al.*, 2014).

The side chain oxygen of Tyr-155 deprotonates the C-4 oxygen of the intermediate (atomic distance is 2.7 Å in NahE-HN) for cleavage of the C-3/C-4 bond of HN, as proposed for NAL, DHDPS and KDGA. The pyruvate enamine adduct on Lys-183 tautomerizes and salicylaldehyde is released. Release of salicylaldhyde likely precedes pyruvate hydrolysis since salicylaldehyde sterically blocks the exit of pyruvate at the opening to the active site, as observed in NahE-SP (**Figure 2-15**) and is consistent with the sequential Uni-Bi mechanism suggested for NAL and DHDPS.



Scheme 2-2 Proposed mechanism for NahE using HBP.

2.4.3 The Mechanism of PhdJ Using CBP

The k_{cat} and K_m values for PhdJ (6.5 ± 0.4 s^{-1} and 79 ± 7 μM) using CBP are similar to those published for the ortholog from *Nocardioides* strain KP7 (55% sequence

identity, 13 s^{-1} and 50 μM, Iwabuchi & Harayama, 1998). The enzyme exhibits a clear preference for CBP, as its k_{cat}/K_M value is 26-fold and 1000-fold greater than that using HBP and BP, respectively.

Crystal structures were obtained for PhdJ and PhdJ in complex with CBP (and products CBA and pyruvate bound in other sites of the same structure), the 4-hydroxy intermediate CN, and CBA (Schiff base). Native PhdJ and PhdJ complex crystal structures are superimposable with RMSD values less than 0.25 Å (**Figure 2-26A,B**). Like NahE, the positioning of side chains in the active site are similar, with the most significant differences being in the small helix that caps the β-barrel.

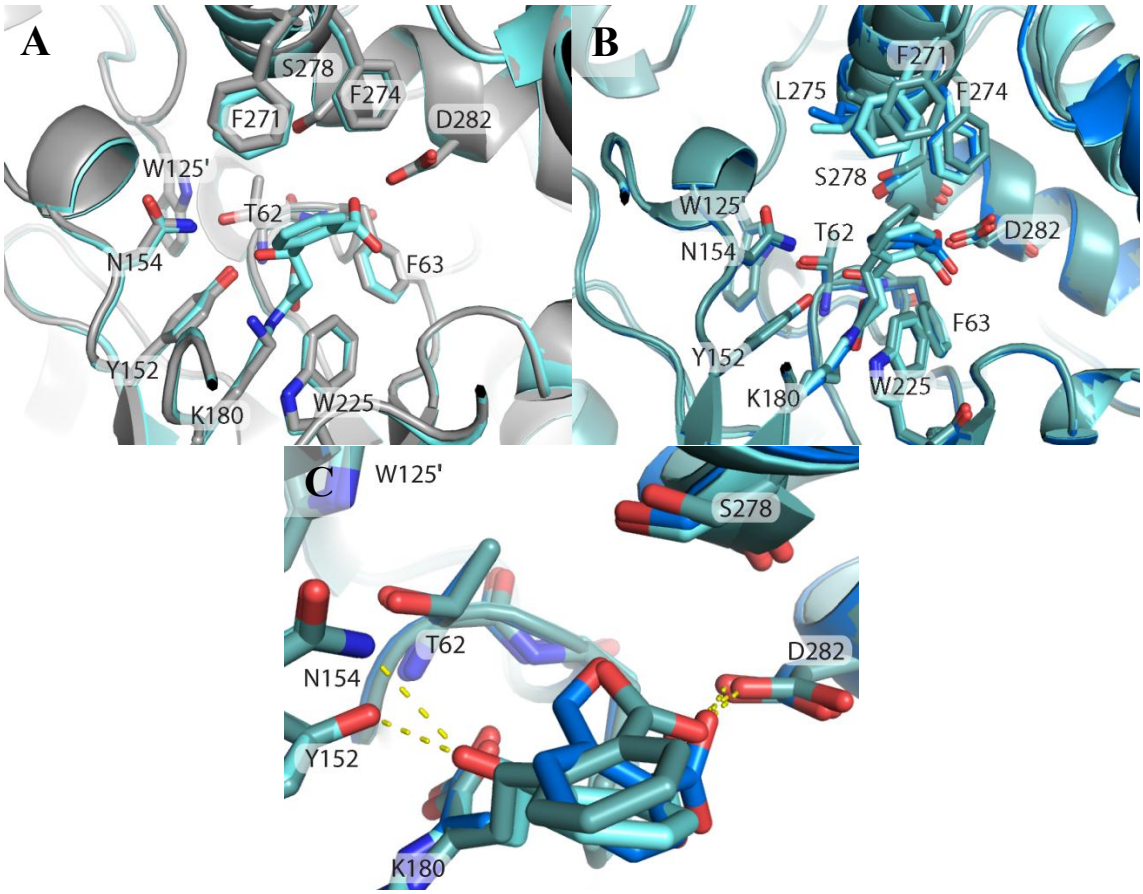


Figure 2-28 Superposition of PhdJ crystal stuctures.

A) Alignment of PhdJ (grey) and PhdJ-CN (aqua marine). **B)** Alignment of PhdJ-CBP (light teal), PhdJ-CN (aqua marine), and PhdJ-SP (marine). **C)** Close-up of Schiff base adducts of substrate CBP (light teal), 4-hydroxy intermediate (CN, aqua marine), and products CBA and pyruvate (marine), shown in **B**.

Water deprotonation via the side chain oxygen of Tyr-152 may occur as was proposed for NahE (**Scheme 2-2**). The oxygen of Tyr-152 is 4.8 Å from W_1 in the PhdJ-CBP, and only 3.8 Å from the benzylic carbon, and is positioned on the *Re* face to which

the water molecule adds. The likelihood of deprotonation of water by the side chain hydroxyl group of Tyr-152 and the C-1 carboxylate of CBP is increased with the understanding that CBP is in a nonproductive orientation, and that CBP likely undergoes rearrangement prior to water addition.

However, the *o*-carboxylate of CBP is 3.0 Å from W_1 in PhdJ-CBP, as shown in **Figure 2-29A**, which indicates could be a general base for water addition. The substrate-assisted catalysis implies PhdJ would use a different mechanism than that proposed for NahE. This seems unlikely given the conserved residues Tyr-152 and Asn-154 (PhdJ numbering) were proposed to activate water for NahE (**Scheme 2-2**) and the 45% identity between the enzymes. However, the possibility of water deprotonation by the *ortho*-carboxylate group of CBP is worth considering due to the proximity of W_1 to the *ortho*-carboxlate in the PhdJ-CBP (3.0 Å). This mechanism for PhdJ using CBP (**Scheme 2-3**) begins with Schiff base formation between the side chain amino group Lys-183 and the C-2 of CBP mediated by the side chain of Tyr-152 as proposed for NahE and other NAL subgroup members. The *o*-carboxylate oxygen of CBP deprotonates water with the assistance of a water molecule and the side chain nitrogen of Asn-154. Water activation by another water molecule stabilized to the side chain hydroxyl group of Ser-278 cannot be ruled out. Another possibility, noted for NahE (**Section 2.4.2**) is the deprotonaton of water via the C-1 carboxylate, which is 5.2 Å from W_1 in PhdJ (though only 3.3 Å from C-4 where water is added in PhdJ-CBP). After water is added, electron rearrangement then occurs to form an enamine intermediate. The protonated *o*-carboxylate then transfers the proton to C-3 (atomic distance of 3.4 Å in PhdJ-CN, **Figure 2-29B**) to complete *anti*-

addition. PhdJ-CN shows that the side chain hydroxyl of Tyr-152 and water molecule is 4.1 and 4.0 Å from C-3 (**Figure 2-27B**), making these groups less likely proton donors to C-3 during hydration.

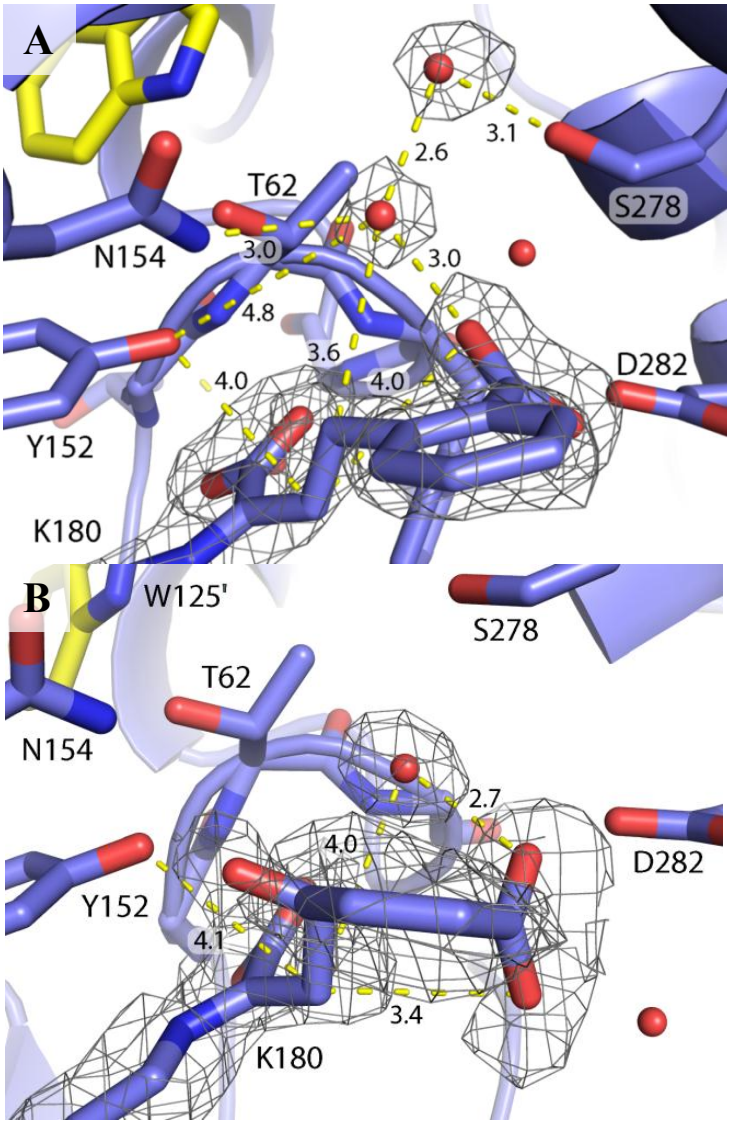
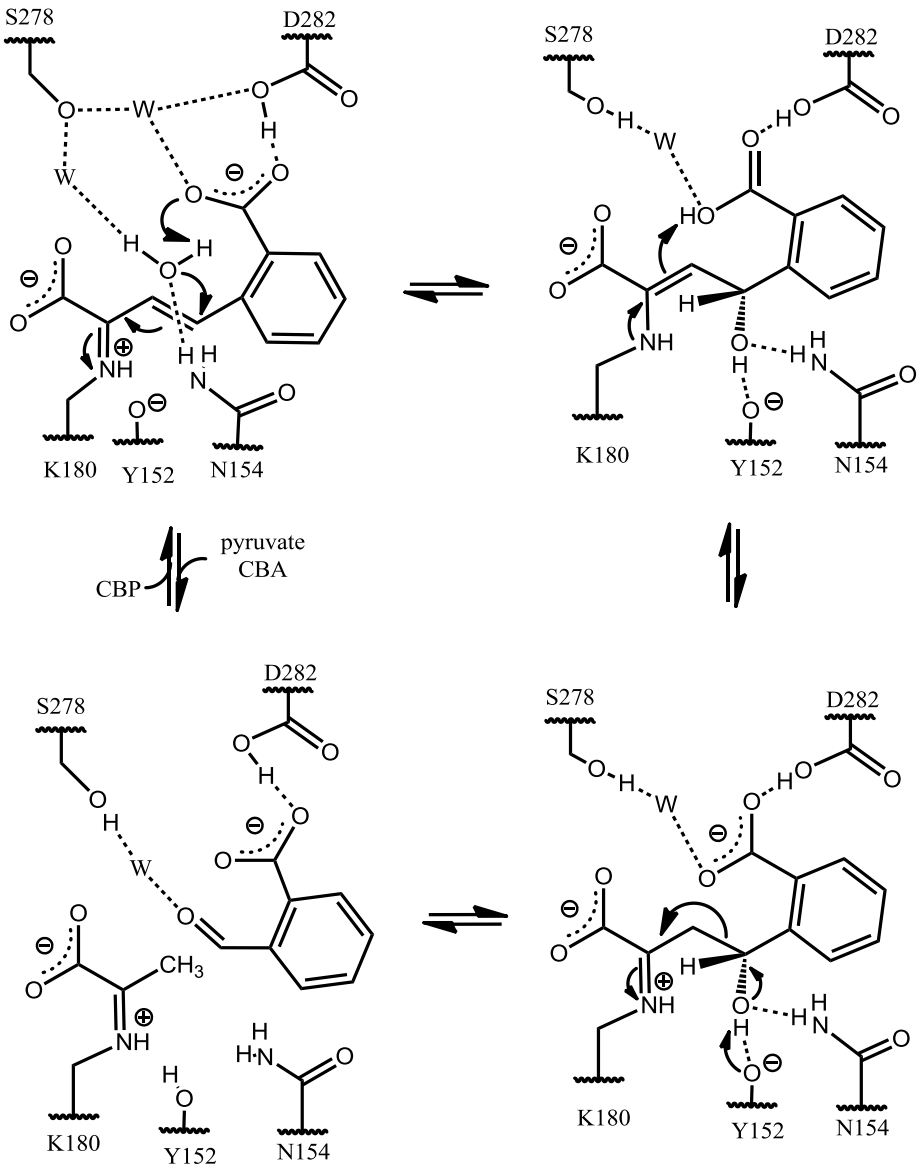


Figure 2-29 Candidates for water activation and proton donation during hydration of CBP by PhdJ.

Distances (yellow dashes) are in Å. $2F_o$-F_c electron density map (1.0 σ) surrounds adducts and nearby water moledules. **A)** PhdJ-CBP active site with groups that potentially activate water and protonate C-3. **B)** PhdJ-CN active site with groups that may protonate C-3.

Based on CBP and CN intermediates in PhdJ-CBP and PhdJ-CN crystal structures, hydration is likely accompanied by a 180 degree rotation of the *o*-carboxylate group around the oxygen that is hydrogen bonded to Asp-282. A 0.7 Å shift of the benzylic carbon is also observed when the PhdJ-CBP and Phd-CN structures are compared. These movements may allow access of activated water to C-2, since W_1 is in plane with the double bond in PhdJ-CBP. In addition, these movements that go with hydration appear to bring one *ortho*-carboxylate oxygen close to C-3 (3.4 Å). Imine-enamine tautomerization may occur via this oxygen as the $2F_o$-F_c electron density (at 1.0 σ) protrudes toward C-3 (**Figure 2-29B**).

Like NAL, DHDPS, and KDGA (**Section 1.4**), the retro aldol reaction proceeds by deprotonation of the C-4 oxygen by the side chain of Tyr-152 (distance in PhdJ-CN is 2.9 Å), where the developing anion is stabilized by the side chain amide nitrogen of Asn-154. The pyruvate enamine tautomerizes and CBA is released from the active site. Pyruvate can then be hydrolyzed and released, consistent with a sequential Uni-Bi reaction.



Scheme 2-3 Proposed mechanism of PhdJ using CBP.

2.4.4 The S278N and D282E PhdJ Mutants

Steady state kinetic parameters were determined for the S278N and D282E PhdJ mutants, which were designed to mimic the active site of NahE (**Figure 2-30**). The K_M values for S278N and D282E PhdJ mutants using HBP are both 2-fold lower than with wild-type, and the k_{cat} values are reduced 54- and 35-fold, respectively. Although the mutation impeded turnover, the parameters are consistent improved binding of HBP. The k_{cat}/K_M values for S278N and D282E PhdJ mutants are close to 10^4-fold less than that for NahE, thus these point mutations do not imitate NahE in its efficiency using HBP. Double mutant S278N D282E PhdJ may have kinetic parameters more similar to wild-type NahE when using HBP, though similar enzyme efficiency may require Arg-288 (NahE numbering), which participates in the proposed mechanism for NahE (**Scheme 2-2**).

The D282E mutation was more detrimental to PhdJ activity using CBP, as the k_{cat}/K_M decreased 1500-fold, whereas the S278N mutant merely decreased only 23-fold. Although the drop for the D282E mutant seems to indicate Asp-282 might activate water, especially with the presedence for carboxylate side chains to do so (Bahnson *et al.*, 2002, Bennett *et al.*, 2008, Poelarends *et al.*, 2003, Huddleston *et al.*, 2014), Glu-282 is 6.5 Å from W_1 in PhdJ-CBP (6.2 Å from C-4, **Figure 2-30**), and its closest side chain oxygen to W_1 interacts with an *ortho*-carboxylate oxygen in each liganded PhdJ structure. The extra methylene on the side chain in the D282E mutant might instead disrupt the interaction of the carboxylate oxygen of Glu-282 with the *o*-carboxylate of CBP.

Surprisingly, the S278N and D282E PhdJ mutants processed BP 2.2-fold and 34-fold more efficiently (k_{cat}/K_M) than wild-type PhdJ. Replacing Asp-282 with glutamate may displace the carboxylate group away from the phenyl ring of bound BP, increasing the affinity of BP and PhdJ.

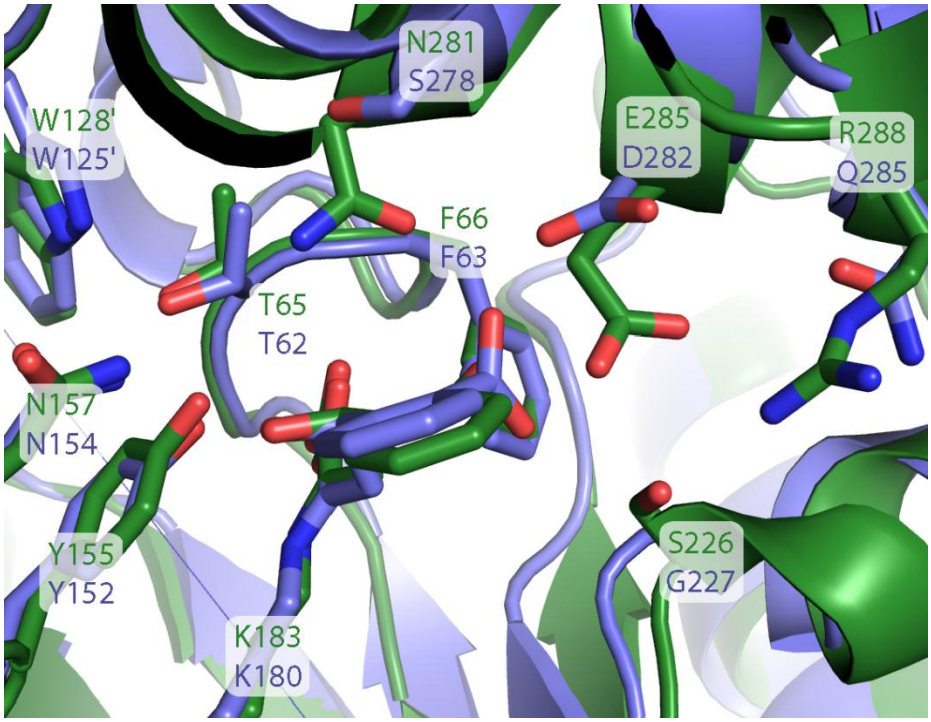


Figure 2-30 Superposition of NahE-HN (green) and PhdJ-CN (slate) active sites.

2.4.5 Structure and Function of PhdG

The proposed role of PhdG in the phenanthrene catabolic pathway could not be confirmed in our hands (**Figure 1-8**). However, PhdG is functionally linked to its paralog PhdJ by its ability to convert HBP to salicylaldehyde and pyruvate, a reaction in the

naphthalene degradation pathway in *M. vanbaalenii* PYR-1. The recently discovered PhdG activity using HBP was assessed by UV/Vis spectrophotometry (**Section 2.3.1**). Both enzymes have the secondary residues conserved in hydratase-aldolases (Asn-157 and Trp-224 in NahE), but tertiary residues (which vary among hydratase-aldolases) Asn-281 and Glu-285 that might be involved in NahE catalysis (**Scheme 2-2**) are replaced by Gln-272 and Trp-276 in PhdG (**Figure 2-32**).

Crystal structures were obtained of native PhdG and PhdG in complex with pyruvate carbinolamine. The carbinolamine adduct observed in the PhdG-Pyr is unexpected, as pyruvoyl adducts in NahE and PhdJ were involved in Schiff base. However, carbinolamine intermediates have been observed before in several type I aldolases (Heine *et al.*, 2001, Lorentzen *et al.*, 2005). The carbinolamine oxygen atom in archaeal fructose-1,6-bisphosphate aldolase hydrogen bonds with the side chain oxygen of Tyr-146 (conserved in PhdG) which was proposed to deprontonate the carbinolamine for Schiff base formation. The rabbit fructose-1,6-bisphosphate aldolase A has Glu-187 at the equivalent position which was suggested to be involved in deprotonation and dehydration of the carbinolamine, based on kinetic studies of Glu-187 mutants. The interaction of C-2 oxygen of the carbinolamine adduct and Tyr-152 in the crystal structure of PhdG-Pyr suggests proton transfer might occur between these groups providing further evidence that the side chain oxygen of Tyr-152 (PhdG numbering) mediates Schiff Base formation for hydratase-aldolases and NAL subgroup members.

Although the overall structures of PhdG and PhdG-Pyr are nearly superimposable (RMSD of 0.18), several differences are observed in the active site. First, the difference

in dihedral angle between Cα and Cβ of Gln-272 is ~180°. In PhdG-Pyr, the difference accompanies a hydrogen bond between the side chain amide nitrogen Gln-272 and a water that interacts with the carbinolamine oxygen of the adduct. Secondly, the side chain of Thr-182 hydrogen bonds with the side chain nitrogen Lys-180 in PhdG, but the side chain oxygen of Thr-182 points away from Lys-180 bound to the carbinolamine adduct in PhdG-Pyr.

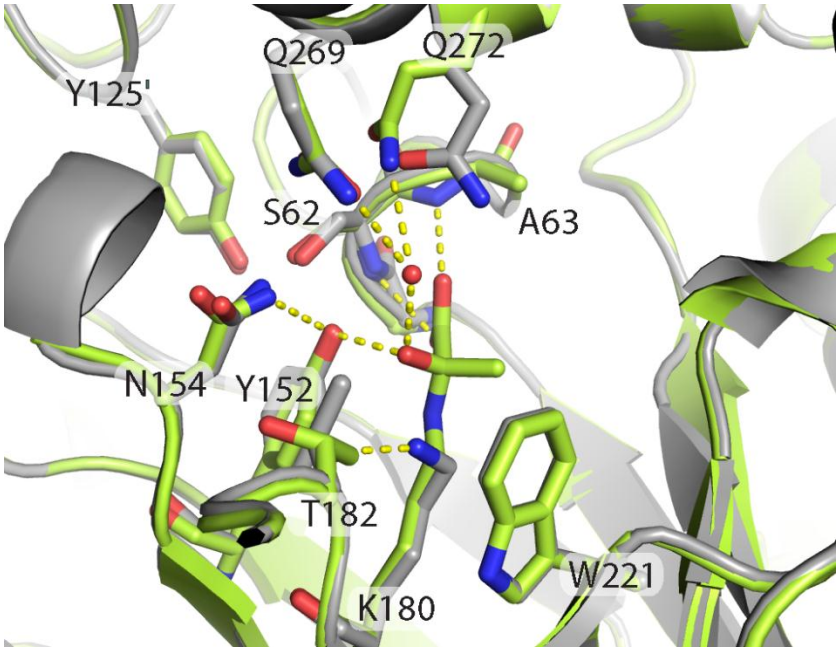


Figure 2-31 Superposition of PhdG and PhdG-Pyr.

The mechanism of PhdG using HBP and BP likely resembles that proposed for NahE using HBP, though Glu-269 and Glu-272 may be involved in water activation or the retro aldol reaction as they extend further towards the pyruvoyl adduct in PhdG-Pyr

than the equivalent residues in NahE (Ser-278 and Asn-281) and PhdJ (Leu-275 and Ser-278) do (**Figure 2-32**). The lack of an equivalent to Asp-282 in PhdJ and presence of nearby Ile-223 likely prevents binding of the CBP, as the interaction between the *o*-carboxylate and the protein in this region may be unfavorable sterically and/or electrostatically.

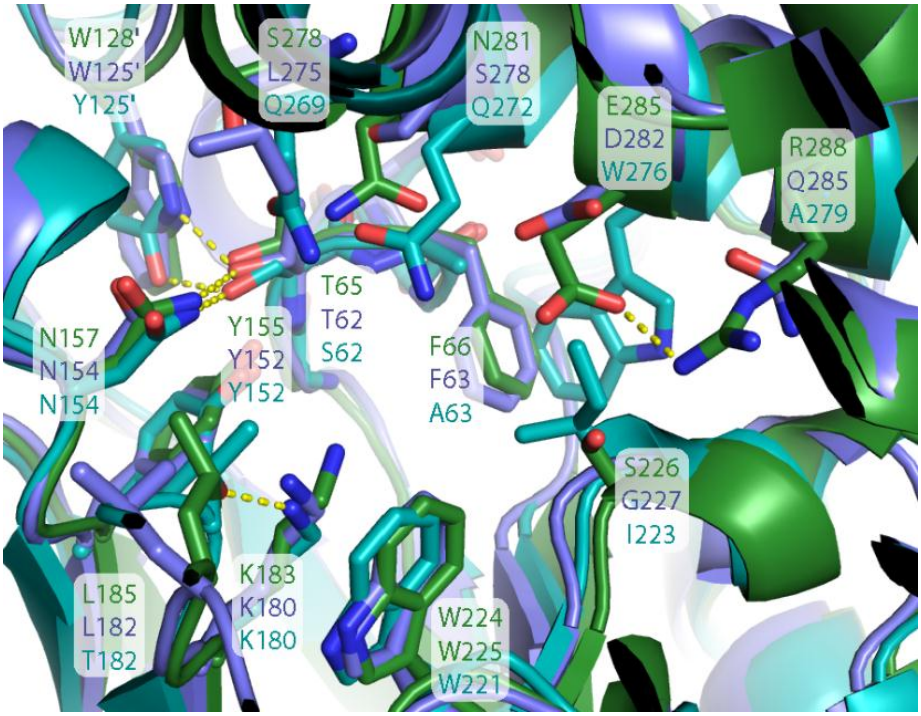


Figure 2-32 Superposition of NahE (forest), PhdJ (slate), and PhdG (light teal).

The k_{cat}/K_M value for PhdG using BP, which is not an intermediate in PAH metabolism in *M. vanbaalenii* PYR-1 is 15- and 110-fold greater than that for NahE or PhdJ. The increased efficiency of PhdG compared to NahE is due to both an increase in

k_{cat} and a decrease in K_M. The difference k_{cat}/K_M value may influenced by Trp-276 in PhdG which replaces Glu-285 in NahE and Asp-282 in PhdJ, **Figure 2-32**).

The K_M value with respect to pyruvate for PhdG using cinnamaldehyde and pyruvate with 22-fold less than that for NahE. Additionally, the K_M value for PhdG using BP is 3.2-fold less than that for NahE. The basis for the tighter binding in PhdG may be related to Tyr-125 which interacts with the adjacent monomer (replaces Trp-128 in NahE), and Gln-272 which hydrogen bonds to the carbinolamine oxygen via a water molecule and is replaced by Asn-281 in NahE (**Figure 2-32**). The formation of the carbinolamine might also be related to the higher affinity of PhdG for α-keto acids.

2.4.6 Biological Roles for NahE, PhdJ, and PhdG

Kinetic and ^{1}H spectrometric data indicate NahE converts HBP, an intermediate in naphthalene degradation (as was known previously, **Section 1.2.3**), and processes CBP, an analoguous metabolite in phenanthrene degradation via 3,4-dihydroxyphenanthrene. The latter activity indicates NahE may be used in phenanthrene metabolism in *Pseudomonas fluorescens* containing an NAH7-like plasmid (Menn *et al.*, 1993), albeit at a ~10^4-fold lower k_{cat}/K_M value than in naphthalene degradation. The low K_M (1.2 μM) of NahE using HBP may allow for the prevention of buildup of HBP and HCCA in the pathway. NahE may control flux through the pathway due to its inhibition by salicylaldehyde, as well as its slow inactivation by salicyaldehyde.

PhdJ is relatively nonspecific compared to NahE, as the k_{cat}/K_M value using CBP, a phenanthrene catabolic intermediate, is only 26-fold greater than when using HBP.

Since PhdG also processes HBP (the kinetic parameters are not determined yet), both enzymes participate in naphthalene catabolism in *M. vanbaalenii* PYR-1, as the bacterium degrades naphthalene through salicylate as is established in *Pseudomonas putida* G7 (Kelley *et al.*, 1990). The lack of activity of PhdG with 1-hydroxy-2-naphthaldehyde and pyruvate (reverse reaction that is proposed to be PhdG-catalyzed, Stingley, *et al.*, 2004) in our laboratory and the redundancy of PhdG in its ability to process HBP suggests it might act on other pathway intermediates in *M. vanbaalenii* PYR-1.

Chapter 3: The 4-Oxalocrotonate Tautomerase, *cis*-3-Chloroacrylic Acid Dehalogenase, and Malonate Semialdehyde Decarboxylase Subgroups of the Tautomerase Superfamily

Characterized members of the tautomerase superfamily (**TSF**) are distinct in that they have a catalytic N-terminal proline. In addition, they have a common fold. Some TSF members are composed of a single β-α-β unit (55-75 amino acids), while others are made up of two β-α-β units connected by a linker (105-150 residues, **Figure 3-1**). Thus far, over 15,000 predicted gene products have been annotated as TSF members, and characterized members are known to catalyze tautomerization, isomerization, dehalogenation, hydration, decarboxylation, and oxygenation reactions (Poelarends *et al.*, 2008, Zandvoort *et al.*, 2012, Zandvoort *et al.*, 2012, Baas *et al.*, 2015). Studies of these enzymes uncover strategies used by nature to generate diversity, and it might be possible to mimic these principles to develop new proline-based enzymes (Poelarends *et al.*, 2008). The TSF consists of five subgroups designated by the first characterized member. They are the 4-oxalocrotonate tautomerase (**4-OT**), *cis*-3-chloroacrylic acid dehalogenase (***cis*-CaaD**), macrophage migration inhibitory factor (**MIF**), malonate semialdehyde decarboxylase (**MSAD**), and 5-(carboxymethyl)-2-hydroxymuconate isomerase (**CHMI**) subgroups. Grouping of enzymes is based on sequence analysis and experimentally characterized functions. The work presented here focuses on the structures and functions of enzymes within the *cis*-CaaD, MSAD, and 4-OT subgroups.

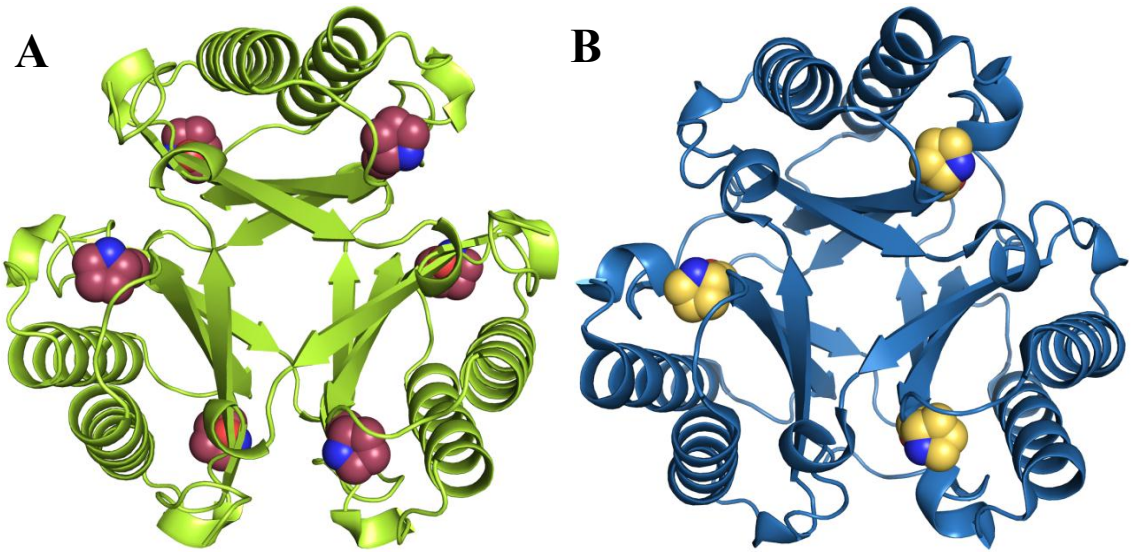


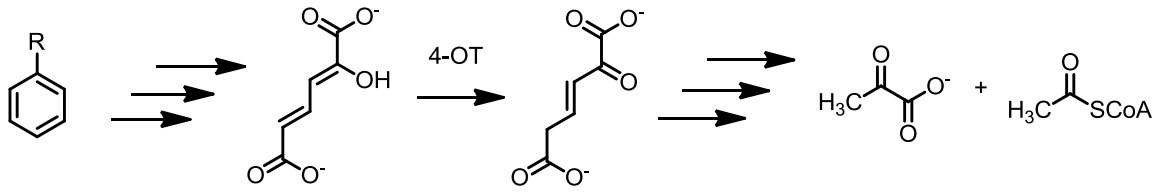
Figure 3-1 Assemblies of long and short monomers in the TSF.

A) Canonical 4-OT (limon) is a hexamer, made up of short monomers that have six Pro-1 residues (violet purple, PDB entry 1OTF). The structure can be thought of as a trimer of dimers. **B)** *cis*-CaaD is made up of long monomers with three Pro-1 residues (yellow orange, PDB entry 3MF8).

3.1 4-OXALOCROTONATE TAUTOMERASE

4-Oxalocrotonate tautomerase (**4-OT**) is a small enzyme in the catabolic meta-fission pathway that converts monocyclic aromatic hydrocarbons to compounds that can be shuttled into the Krebs cycle. 4-OT is encoded by the *xylH* gene on the TOL plasmid WW0, which is similar to the NAH7 plasmid (**Section 1.2**) in *Pseudomonas putida* mt-1. The hexameric enzyme catalyzes the tautomerization of 2-hydroxymuconate (**2-HM**) to 2-keto-3-hexenedioate shown in **Scheme 3-1** (Whitman *et al.*, 1991). Each chain has 62

amino acids and folds into a single β-α-β motif with a C-terminal β-hairpin (**Figure 3-2A**). The latter feature is important in forming the hexamer (Whitman, 2002).



Scheme 3-1 Role of 4-OT in the meta-fission pathway.

Pro-1, Arg-11, Arg-39, and Phe-50 are crucial in the tautomerization reaction, and their roles have been studied for over 25 years by NMR, crystallography, mutagenesis, kinetic, and inhibition studies (**Figure 3-2B**). Pro-1 has a pK_a of about 6.4, which is lowered more than 3 units from the model compound proline amide, due in part to the nearby hydrophobic Phe-50′ (Czerwinski *et al.*, 2001). The prime symbol indicates that a residue comes from the other monomer in the dimer and the double prime indicates that a residue comes from an adjacent dimer. The hexameric structure is a trimer of dimers (**Figure 3-1A**), and residues from three adjacent subunits participate in catalysis. For one active site containing Pro-1, the proximal Arg-11′ and Phe-50′ are from the other subunit in same dimer, and Arg-39″ is from a neighboring dimer.

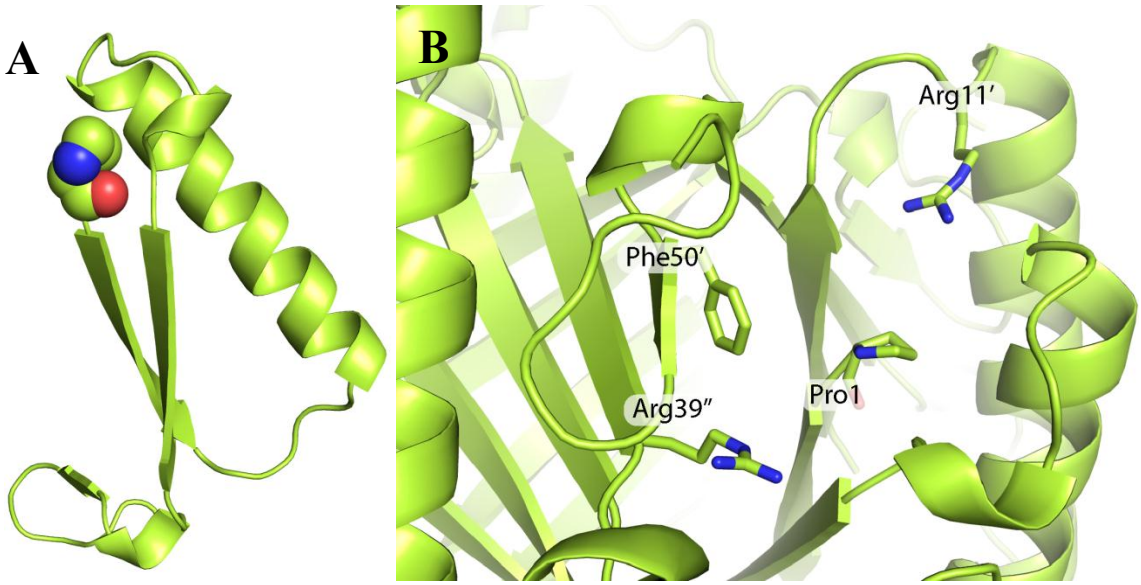
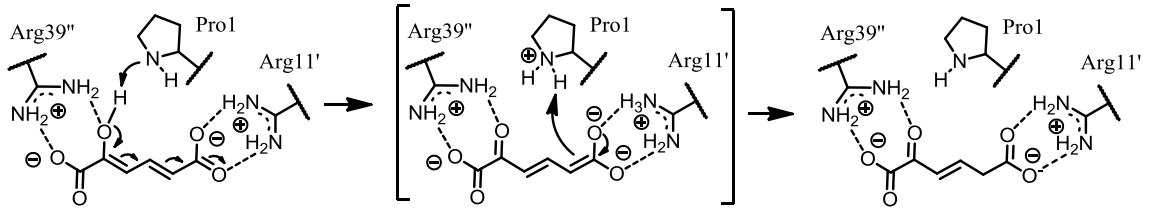


Figure 3-2 Crystal structure of 4-OT (1.9 Å resolution, PDB entry 1OTF).

A) 4-OT monomer with Pro-1 shown as a sphere. **B)** Key active site residues shown in hexameric assembly.

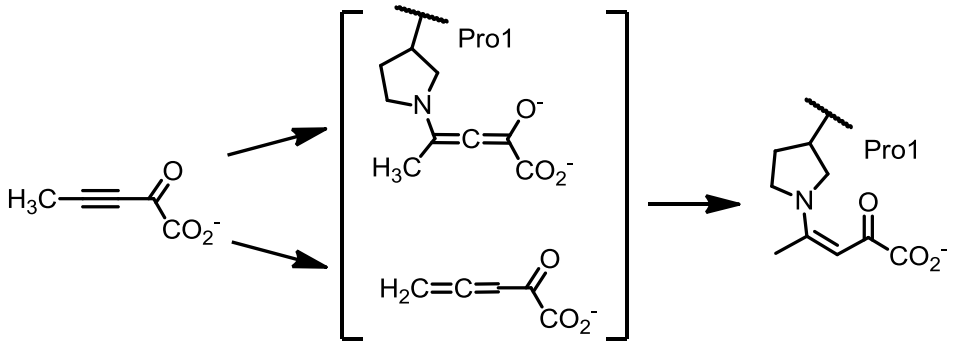
The low pK_a value allows Pro-1 to function as a general base catalyst for deprotonation of the 2-hydroxyl group of 2-HM and rearrangement to form an enediolate at C-5 and C-6 (**Scheme 3-2**). Arg39″ is involved in catalysis as the side chain stabilizes the developing anion at the 2-hydroxyl group during deprotonation. Arg-11′ forms a salt bridge with the C-6 carboxylate and is thought to polarize 2-HM so that electrons are pulled towards C-6 for protonation at the C-5 position (Czerwinski *et al.*, 1999, Harris *et al.*, 1999). Delivery of a proton to C-5 of the enediolate intermediate completes the tautomerization (Stivers *et al.*, 1996, Czerwinski *et al.*, 1997, Poelarends *et al.*, 2008). The protonation is stereoselective and the 5(*S*)-isomer is generated when the reaction is performed in D_2O (Whitman *et al.*, 1992).



Scheme 3-2 Mechanism of 2-HM tautomerization by 4-OT.

2-Oxo-3-pentynoate (**2-OP**) is an acetylene compound initially discovered to be an irreversible inhibitor of 4-OT, initiated by attack of the prolyl nitrogen (Pro-1) at C-4. Inactivation as a function of 2-OP concentration suggested half-of-the-sites stoichiometry (Johnson *et al.*, 1997), though the crystal strcture of 4-OT in complex with 2-OP revealed that each Pro-1 residue was modified (Taylor *et al.*, 1998). The half-of-the-sites stoichiometry observed in solution was proposed to be a result of allostery, where inactivation of the first monomer precludes the inactivation in the second monomer of one dimer. Inactivation may proceed through initial Micheal addition to form an allene adduct (top intermediate, **Scheme 3-3**), followed by protonation at C-3. Alternatively, 4-OT may catalyze allylic rearrangement of 2-OP by proton abstraction at C-5 to form an allene (bottom intermediate, **Scheme 3-3**), which is poised for attack by Pro-1 at C-4 and protonation at C-5. In the first mechanism, 2-OP is an active site-directed irreversible inhibitor. In the second mechanism, where rearrangement occurs first, 2-OP is a mechanism-based inhibitor. The interactions observed in the crystal structure of the inactivated enzyme suggested that Arg-39″ and an ordered water polarize of the 2-oxo

group of 2-OP to assist Micheal addition of Pro-1 (Johnson *et al.*, 1997, Taylor *et al.*, 1998). Additionally, Arg-61′, the C-terminal residue of 4-OT, was coordinated to the C-5 carboxylate in the inactivated structure. These interactions may parallel those that occur in the tautomerization reaction with 2-HM except for the interaction with Arg-61′, which was subsequently shown by mutagenesis not to be inolved in the tautomerization reaction.

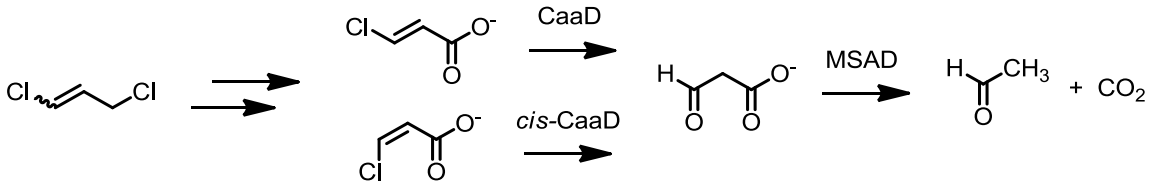


Scheme 3-3 Inactivation of 4-OT by 2-OP.

3.2 1,3-DICHLOROPROPENE CATABOLIC PATHWAY

The synthetic compound 1,3-dichloropropene was introduced into the environment in the 1940s and is used as a nematocide (Poelarends *et al.*, 1998). The compound is rapidly degraded in the soil in part due to the 1,3-dichloropropene catabolic pathway. This pathway converts racemic 1,3-dichloropropene to acetaldehyde and CO_2 (**Scheme 3-4**) and involves three steps to form *cis*- and *trans*-3-chloroacrylate (Poelarends & Whitman, 2004). The *cis*- and *trans*- isomers are processed by isomer-

specific *cis*-CaaD (Poelarends *et al.*, 2003) and CaaD (Poelarends *et al.*, 2001), respectively, to produce malonate semialdehyde. Malonate semialdehyde decarboxylase (MSAD), the final enzyme in the pathway, is a TSF member that catalyzes the decarboxylation of malonate semialdehyde (**Section 3.5**). Subsequently, the product acetaldehyde is likely channeled to the Krebs cycle. The dehalogenation reactions involving CaaD and *cis*-CaaD, and the decarboxylation reaction catalyzed by MSAD are cofactor-independent.

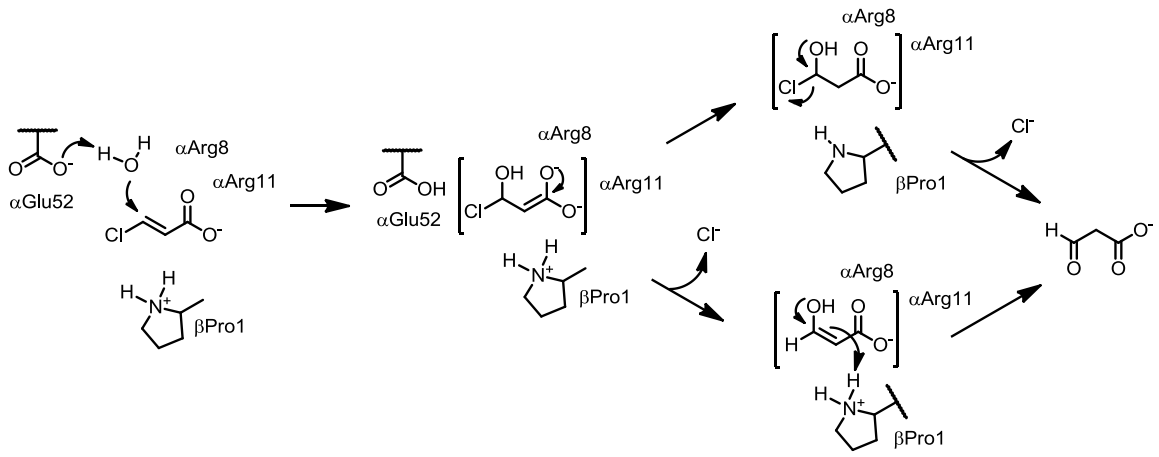


Scheme 3-4 Degradation of racemic 1,3-dichloropropene.

3.3 *TRANS*-3-CHLOROACRYLIC ACID DEHALOGENASE

1,3-Dichloropropene is converted by three enzymes to form a mixture of *cis*- and *trans*-3-chloroacrylate. The *trans* isomer is converted to malonate semialdehyde by *trans*-3-chloroacrylic acid dehalogenase (CaaD), through hydrolytic cleavage of the β-vinylic carbon-chlorine bond (**Scheme 3-5**). CaaD is a heterohexameric protein composed of 3 α-subunits and 3 β-subunits, where each monomer is composed of a single β-α-β motif with Pro-1 at the N-terminus (Poelarends *et al.*, 2001). The mechanism for CaaD involves

αGlu-52, αArg-8 and αArg-11, where αGlu-52 activates water for addition at C-3. The two arginine residues αArg-8 and αArg-11 stabilize the C-1 carboxylate and increase electrophilicity at C-3. The arginine residues might stabilize a enediolate species which forms after hydration. Protonation could occur at C-2 (**Scheme 3-5**, upper pathway) which is followed by aldehyde formation and expulsion of chloride. In an alternate mechanism, electrons are pushed back to C-3 (**Scheme 3-5**, lower pathway) concomitant with chloride release. Protonation can subsequently occur at C-2 during tautomerization of the enolate in the formation of malonate semialdehyde.



Scheme 3-5 Mechanism for CaaD using *trans*-3-chloroacrylate.

CaaD was the first TSF member identified that uses Pro-1 as an general acid. In addition to dehalogenation of *trans*-3-chloroacrylate, CaaD catalyzes the hydration of 2-OP to form acetopyruvate (**Scheme 3-7**) and is inactivated by 3-halopropiolates (Wang *et al.*, 2003). These reactions are discussed in more detail in the next section.

3.4 *CIS*-3-CHLOROACRYLIC ACID DEHALOGENASE

cis-3-Chloroacrylic acid dehalogenase (*cis*-CaaD) is the most thoroughly studied enzyme within the *cis*-CaaD subgroup of the TSF. *cis*-CaaD was isolated from *Coryneform* bacterium strain FG41, which is capable of using 1,3-dichloropropene as its sole carbon source (van Hylckama Vlieg & Janssen, 1991). *cis*-CaaD is a trimer where each member has 149 amino acids that codes two β-α-β units (**Figure 3-3A**). There are three active sites where each has a Pro-1. Sequence analysis and mutagenesis showed that Pro-1, Arg-70, Arg-73, and Glu-114 are important residues for *cis*-CaaD catalysis (**Figure 3-3B**) (Poelarends *et al.*, 2004). These residues are conserved in *trans*-3-chloroacrylic acid dehalogenase (CaaD), and dehalogenation of *cis*-3-chloroacrylate by *cis*-CaaD was initially assumed to proceed by a similar mechanism.

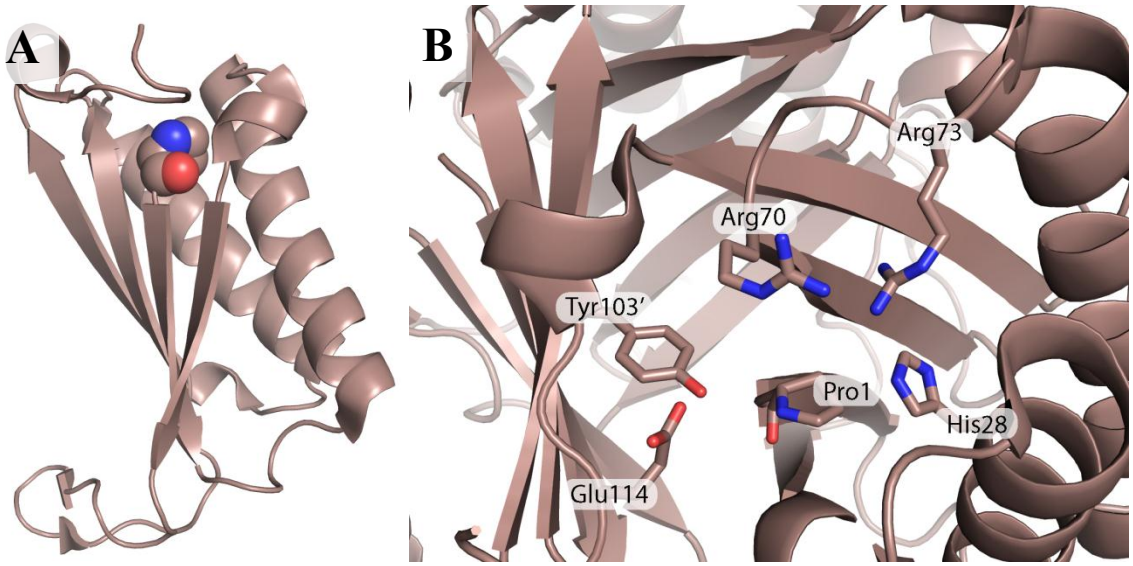
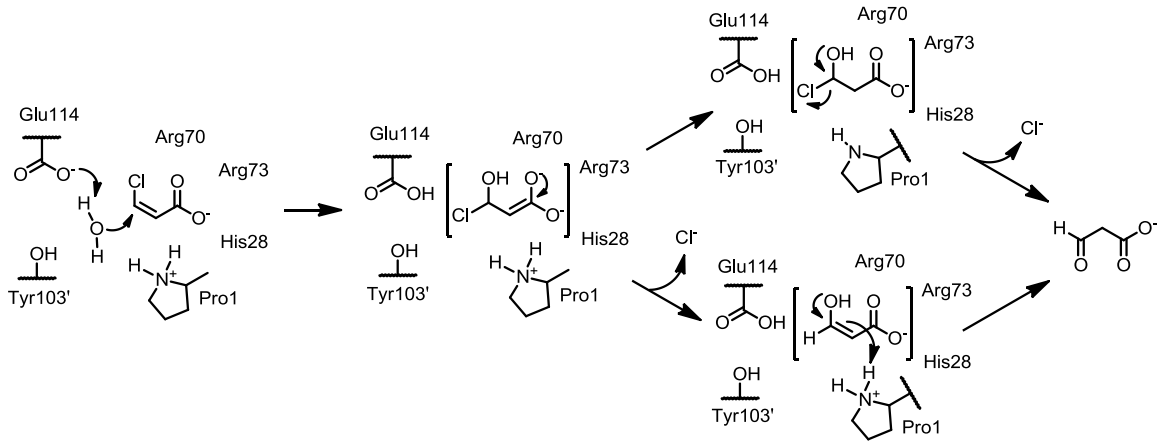


Figure 3-3 Crystal structure of Native *cis*-CaaD (2.0 Å resolution, PDB entry 3MF8) **A)** *cis*-CaaD monomer with Pro-1 shown as a sphere. **B)** Active site of trimeric *cis*-CaaD.

pH rate profiles suggest that the pK_a of Pro-1 in *cis*-CaaD is ~9.3 (Poelarends *et al.*, 2008), which indicates that Pro-1 is a general acid in the dehalogenation of *cis*-3-chloroacrylic acid (**Scheme 3-6**). A crystal structure of *cis*-CaaD inactivated by the substrate analogue (*R*)-oxirane-2-carboxylate indicates that Arg-70 and Arg-73 polarize *cis*-3-chloroacrylate prior to hydration (Poelarends *et al.*, 2004, de Jong *et al.*, 2007). Glu-114 was the obvious candidate for water activation since the equivalent residue for CaaD was required for activity; the αE52Q CaaD variant is completely inactive (de Jong *et al.*, 2007). The process for water activation is different for *cis*-CaaD, though, as the k_{cat}/K_M value for the E114Q mutant decreased only 10-fold (Poelarends *et al.*, 2003).

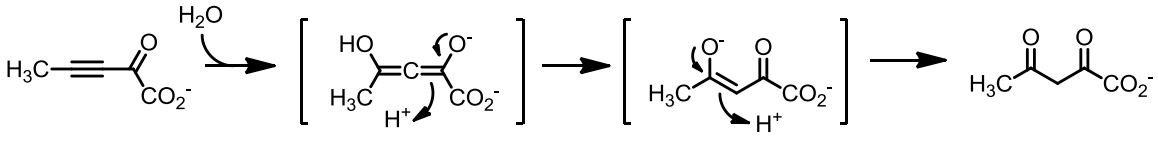
The *cis*-CaaD crystal structure modified by (*R*)-oxirane-2-carboxylate implicated Tyr-103′ as a residue that assists Glu-114 in water activation (de Jong *et al.*, 2007).

Although the k_{cat}/K_M value decreased slightly for the Y103F variant, the k_{cat} dropped more than tenfold, and the activity of the Y103F/E114Q variant was severely compromised when the reaction was followed by ^{1}H NMR spectroscopy (de Jong *et al.*, 2007). Tyr-103′ might position the water molecule for activation or it might position Glu-114 for water activation, or the mechanism might involve some combination of the two functions. Tyr-103′ also provides specificity as it contributes to the formation of a U-shaped active site, which differs from the CaaD active site cavity. Tyr-103′ could block the chloro group of *trans*-3-chloroacrylate and prevent binding. The *cis*-CaaD structure inactivated by (*R*)-oxirane-2-carboxylate implicated His-28 as another important residue for catalysis. His-28 might accompany Arg-70/Arg-73 in their interaction with the C-1 carboxylate. In addition, the crystal strcture suggested Thr-34, Leu-38, and Leu-119 "cradle" *cis*-3-chloroacrylate in the binding pocket (Poelarends *et al.*, 2008).



Scheme 3-6 Mechanism for *cis*-CaaD using *cis*-3-chloroacrylate.

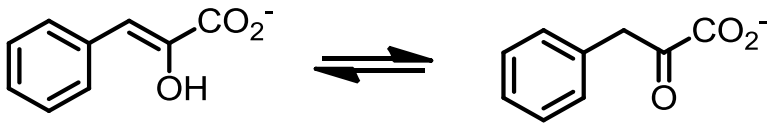
Consistent with the higher pK_a of Pro-1, *cis*-CaaD (like CaaD) catalyzes the hydration of 2-oxo-3-pentynoate (2-OP) shown in **Scheme 3-7** (Poelarends *et al.*, 2003). *cis*-CaaD was inactivated at high 2-OP concentrations (60 mM), and the hydration activity with *cis*-CaaD is 500-fold less efficient than that of CaaD. The hydration likely involves Micheal addition and formation of an allene which then ketonizes. Hydration of 2-OP by TSF members could indicate they exhibit a hydratase activity in their native organism.



Scheme 3-7 Hydration mechanism for 2-OP.

The phenypyruvate tautomerase (PPT) activity (**Scheme 3-8**) of *cis*-CaaD links the enzyme to 4-OT and provides insight into *cis*-3-chloroacrylate dehalogenation. Although the physiological significance of phenylenolpyruvate tautomerization is not known, comparable PPT activity of 4-OT and *cis*-CaaD (in addition so sequence identity) suggests a functional relationship between the two proteins (Poelarends *et al.*, 2007, Poelarends *et al.*, 2008, Burks *et al.*, 2010). Heterohexameric 4-OT from *Chloroflexus aurantiacus* J-10-fl is composed of α and β subunits where each form a dimer (like CaaD) in a trimer of dimers. Studies of PPT activity in this homologue suggest that Pro-1 is the

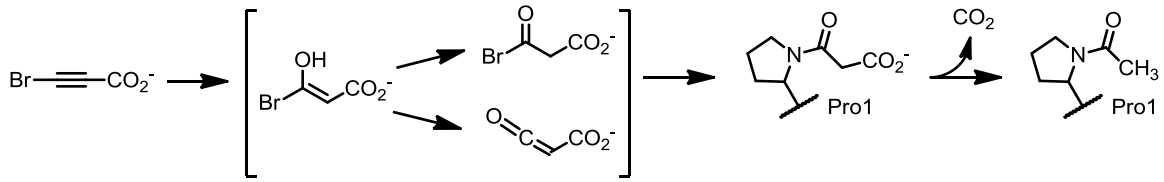
catalytic base and αArg-12 (equivalent to Arg73 of *cis*-CaaD) interacts with the carboxylate of phenylenolpyruvate to increase the efficiency of ketonization. αTrp-51 (equivalent to Met-112 in *cis*-CaaD) might be important in binding the phenyl ring (Burks *et al.*, 2010). The PPT activity suggests that *cis*-CaaD (as well as 4-OT) is capable of ketonizing enol compounds. This tautomerization ability of *cis*-CaaD supports the α,β-elimination mechanism with *cis*-3-chloroacrylic acid (lower pathway in **Scheme 3-6**) and formation of the enol intermediate (Poelarends *et al.*, 2007).



Scheme 3-8 Ketonization of phenylenolpyruvate.

Fragmentation analysis of *cis*-CaaD inactivated after being incubated with 3-bromopropiolate using MALDI-MS showed a malonyl adduct covalently bound to Pro-1 (Poelarends *et al.*, 2003). 3-Bromopropiolate is considered a mechansim-based inhibitor where hydration of 3-bromopropiolate likely occurs (using Glu-114) prior to inactivation. In a possible mechanism proposed for *cis*-CaaD, water addition to 3-bromoproiolate forms an unstable enol species (**Scheme 3-9**). Tautomerization forms an acyl bromide, or bromide is expeled to form a ketene species. Hydration of 3-bromopriololate deprotonates Pro-1, making the residue nucleophilic and poised for attack of the acyl bromide or ketene, producing a malonate adduct (de Jong *et al.*, 2004). Decarboxylation

of the adduct then occurs to form an acetyl modification at Pro-1. The decarboxylation might result from the conditions of MALDI-MS analysis.



Scheme 3-9 Covalent labeling of Pro-1 (prolyl nitrogen) after incubation with 3-bromopropiolate.

3.5 MALONATE SEMIALDEHYDE DECARBOXYLASE

Following dehalogenation reactions in the 1,3-dichloropropene pathway, malonate semialdehyde decarboxylase (MSAD) catalyzes the decarboxylation of malonate semialdehyde. MSAD from *Pseudomonas pavonaceae* 170 is a trimer where each long monomer (129 amino acids) encodes two β-α-β units (**Figure 3-4A**).

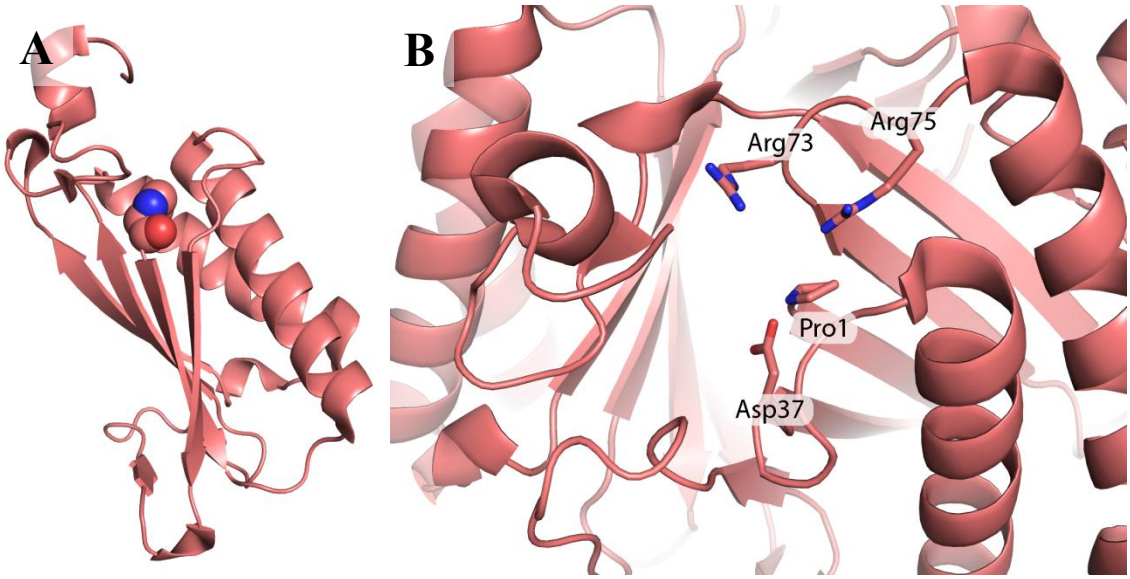
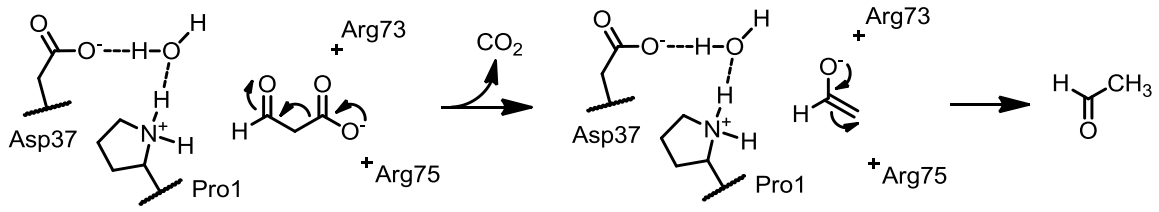


Figure 3-4 Crystal structure of MSAD (1.9 Å resolution, PDB entry 2AAG)**.**

A) MSAD monomer with Pro-1 shown as spheres. **B)** Key active site residues in trimeric MSAD.

Key active site residues include Pro-1, Asp-37, Arg-73, and Arg-75 (**Figure 3-4B**). Pro-1 and Arg-75 are believed to be catalytic because the k_{cat} values for R73A and R75A mutants are reduced more than 100-fold (Poelarends *et al.*, 2004). Pro-1 of MSAD has a high pK_a value of ~9.2, as determined by ^{15}N NMR spectroscopy (Poelarends *et al.*, 2004, Poelarends *et al.*, 2008). Arg-75 is proposed to orient the C-1 carboxylate of malonate semialdhyde favorably for decarboxylation (**Scheme 3-10**) (Poelarends *et al.*, 2005). Arg-73 and Asp-37 are also shown to be important, as the R73A and D37A variants had 1.8% and 0.5% specific activity relative to wild-type (Almrud *et al.*, 2005). A crystal structure of MSAD indicated that a water molecule positioned by Pro-1 and

Asp-37 might polarize the aldehyde oxygen and promote decarboxylation. The active site also contains hydrophobic residues Trp-114, Phe-116, Phe-123, and Leu-128, which could facilitate decarboxylation.



Scheme 3-10 Mechanism for decarboxylation of malonate semialdehyde by MSAD.

The NMR-derived pK_a value of 9.2 for MSAD (Poelarends *et al.*, 2005) is consistent with the observed 2-OP hydration activity (**Scheme 3-7**), a physical property that links the enzyme to CaaD and *cis*-CaaD. Hydration was shown to take place in the same active site as decarboxylation since the P1A, D37A, R73A, and R75A mutants all eliminated activity (Huddleston *et al.*, 2014). Mutagensis and crystallography suggest a mechanism involving polarization of 2-OP by Arg-73 and Arg-75 through interaction with the C-1 carboxylate to increase electophilicity at C-4 and allow hydration. Water is activated by Asp-37 for Micheal addition at C-4. Pro-1 subsequently provides a proton to complete the addition.

Kinetic, mass spectrometric, and crystallographic data indicate MSAD is inactivated by 3-bromo- and 3-chloropropiolates and provides insight into the MSAD-

catalyzed decarboxylation (Almrud *et al.*, 2005, Poelarends *et al.*, 2005). Inactivation proceeds via a acyl halide or ketene prior to labeling at Pro-1 (**Scheme 3-9**). The positioning of the resulting 3-oxopropanoate adduct in the crystal structure suggests that Asp-37 and Arg-75 are involved in both decarboxylation and hydration activities, consistent with the mutagenesis results (Almrud *et al.*, 2005).

Chapter 4: Enzymes Linking 4-Oxalocrotonate Tautomerase and *cis*-3-Chloroacrylic Acid Dehalogenase Subgroups

4.1 INTRODUCTION

Members of the 4-oxalocrotonate tautomerase (**4-OT**) subgroup within the tautomerase superfamily (**TSF**) are composed of short monomers (55-75 amino acids) with a single β-α-β fold that typically oligomerize into a functional hexamer. The oligomeric state of members of the *cis*-3-chloroacrylic acid dehalogenase (***cis*-CaaD**) subgroup is different. They are trimers composed of long monomers (105-150 amino acids) with two β-α-β motifs. Members of the 4-OT subgroup catalyze the tautomerization of 2-hydroxymuconate (**2-HM**), a reaction in the meta-fission pathway for xylenes. Though *cis*-CaaD (**Section 3.4**) catalyzes a dehalogenation reaction in the 1,3-dichloropropene pathway that may have evolved in the last century (**Section 3.2**), other members of the subgroup have low level *cis*-CaaD activity (Poelarends *et al.*, 2008, Baas *et al.*, 2011). *cis*-CaaD subgroup members are believed to have originated through a gene duplication and gene fusion event of an ancestral 4-OT-like enzyme. Closer inspection of the *cis*-CaaD and 4-OT subgroups led to the discovery of three long monomers having sequence homology with both 4-OT and *cis*-CaaD, which may be functional intermediates between the two subgroups.

These "linkers" were identified in the genomes of *Pseudomonas* sp. UW4 (protein designated **Ps01740**, Uniprot code KNIA5), *Pusillimonas sp.* (strain T7-7) (protein

designated **Pt0534**, Uniprot code F4GMX9), and *Burkholderia lata* (strain ATCC 17760) (protein designated **Fused 4-OT**, Uniprot code Q392K7) using a sequence-similarity network technology developed in collaboration with Dr. Patricia Babbitt at The University of California, San Francisco (Davidson *et al., In Preparation*). Of all the ~15,000 proteins that comprise the contemporary TSF, these three linker proteins afford the best scoring connection based on sequence identity between the 4-OT and *cis*-CaaD subgroups (**Table 4-1**). Based on sequence analysis, Ps01740 is closely related to the *cis*-CaaD subgroup, and both Pt0534 and Fused 4-OT are members of the 4-OT subgroup. Crystal structures of these enzymes were determined in order to understand relationships between sequence and structure of enzymes that bridge the two subgroups within the TSF.

Table 4-1 Sequence identity of 4-OT and *cis*-CaaD with Ps01740, Pt0534, and Fused 4-OT.

	cis-CaaD	4-OT and first β-α-β unit[a]	4-OT and second β-α-β unit[a]
Ps01740	24.2	25.4	16.1
Pt0534	18.9	37.1	37.7
Fused 4-OT	16.9	34.9	41.9

[a]The single β-α-β unit of 4-OT is aligned to each β-α-β unit of the linkers.

4.2 EXPERIMENTAL PROCEDURES

Crystallization of Ps01740 and Soaking Procedure The expression and purification of Ps01740, Pt0534, and Fused 4-OT are described previously (LeVieux, *et al., In Preparation*). Crystallization and soaking of crystals were carried out at ~20°C.The crystal used to obtain diffraction data for native Ps01740 (**Ps01740**) and Ps01740 in complex with 2-OP (**Ps01740-2OP**) were obtained in 0.1 M CHES buffer (pH 9.0) and 16-18% PEG 8000. The crystal was soaked for 4 min in a solution of 2-OP (6 mM made in mother liquor) before cryoprotection by 20% glycerol in mother liquor prior to vitrification in liquid nitrogen.

Crystallization of Pt0534 and Soaking with 2-OP. The crystals used to determine Pt0534 in complex with 2-OP (**Pt0534-2OP**) were obtained in 0.2 M ammonium acetate, 0.1 M trisodium citrate (pH 5.6) 24% PEG 4000. The crystals used in the diffraction experiment were grown overnight at 20° C and soaked in 2-OP (6 mM) for 5 min was prior to cryoprotection by 20% glycerol in mother liquor and then vitrified in liquid nitrogen.

Crystallization and Soaking of Fused 4-OT with 2-Hydroxymuconate (2-HM), 2-OP, and Phenylpyruvate. The diffraction-quality crystals for native Fused 4-OT (**F4OT**), and Fused 4-OT in complex with 2-HM, 2-OP, and phenylpyruvate (**F4OT-2HM**, **F4OT-2OP**, and **F4OT-PP**) were obtained in 0.1 M HEPES (pH 8.0) and 28% PEG 8000, whereas the native structure and F4OT-PP were obtained in 0.1 M HEPES (pH 7.5) and 33% PEG 8000. Crystal soaking for F4OT-2HM was performed in mother liquor

with 6 mM 2-HM for 5 min. F4OT-PP structure was determined using crystals soaked with 5 mM phenylpyruvate in mother liquor for 30 min. F4OT-2OP structure was obtained using crystals soaked with 6 mM 2-OP in mother liquor for 11 min. Crystals of all Ps01740, Pt0534, and Fused 4-OT were cryoprotected by 20-25% glycerol in mother liquor.

Data Collection and Structure Determination. X-ray diffraction data for Ps01740, Ps01740-2OP, Pt0534-2OP, F4OT-2HM, and F4OT-2OP were collected at Advanced Light Source (ALS) beamline 5.0.3 (Lawrence Berkeley National Laboratory, Berkeley, CA). Data for F4OT and F4OT-PP structures were collected at ALS beamline 5.0.1. Statistics for data collection are summarized in **Table 4-2, 4-3, 4-4,** and **4-5**. Data were indexed, integrated, and scaled using HKL-2000 (Otwinowski & Minor, 1997). All structures were solved via molecular replacement using Phaser (McCoy, 2007) and Autobuild (Terwilliger *et al.*, 2008) from the PHENIX suite (Adams et al., 2004).

The Cg10062 monomer crystal structure (31.3% sequence identity, PDB code 3N4G) was used as a search model for Ps01740. The solution of Ps01740-2OP was found via molecular replacement using a monomer of Ps01740 as a search model. The structure of Pt0534-2OP was solved via molecular replacement using a 1.2 Å resolution structure of Pt0534 which lacked a C-terminal β-hairpin. The high resolution Pt0534 structure was solved via molecular replacment using a single monomer of Cg10062 (29.0% sequence identity, PDB code 3N4G). All Fused 4-OT structures were solved using a 1.78 Å resolution native structure of Fused 4-OT as a search model.

Refinement of all Ps01740, Pt0534, and Fused 4-OT structures was performed using Phenix Refine (Afonine *et al.*, 2012) Manual model building was done using COOT (Emsley *et al.*, 2010). Model structures were evaluated during and after refinement using Molprobity (Chen *et al.*, 2010).

4.3 RESULTS

4.3.1 Crystal Structure of Ps01740 and Ps01740-2OP

Ps01740 crystallized in the *C*2 space group with 3 monomers per asymmetric unit. The crystal structure of native Ps01740 (**Ps01740**) was determined at a resolution of 2.6 Å via molecular replacement using Cg10062 as a search model. Of the 127 amino acids in Ps01740, the N-terminal 112 residues of each chain were resolved. Each Ps01740 monomer assumes two β-α-β units, the signature fold for members of the TSF. The monomer is comprised of two β-α-β units, a linker connecting the two units, and a C-terminus (**Figure 4-1A**). The β-α-β units form a pair of parallel β strands which interact in an antiparallel manner to make a β-sheet.

The Ps01740 trimer contains a three-fold axis of symmetry (**Figure 4-1B**). The inner core of the Ps01740 trimer is composed of β-strands and is surrounded by an outer wall of alpha helices. The β-hairpins in the linkers and short segments of the C-terminus (residues 109-112) form short anti-parallel β-sheets with the neighboring monomer.

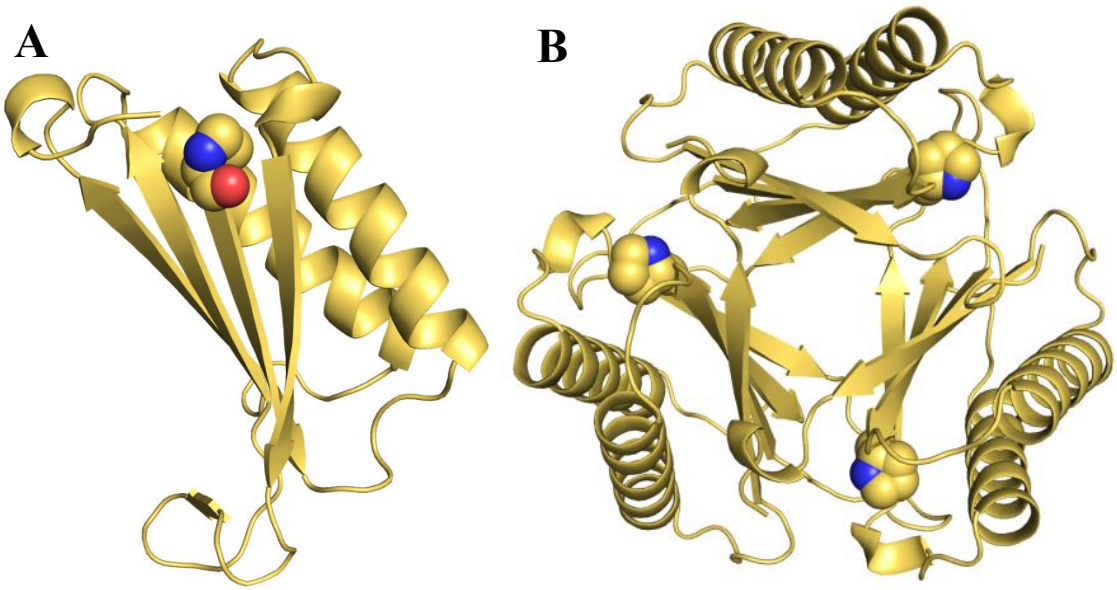


Figure 4-1 The 2.6 Å resolution crystal structure of Ps01740.

A) The Ps01740 monomer with Pro-1 atoms shown as spheres. **B)** The Ps01740 hexamer.

The active site of Ps01740 contains Pro-1, His-28, Arg-68, Arg-71, and Glu-112, which mirror active site residues in *cis*-CaaD (**Figure 4-2A**). However, one catalytic residue in *cis*-CaaD, Tyr-103′ (*cis*-CaaD numbering) is missing in Ps01740, and is substituted with Ala-101′. Glu-112 is the C-terminal residue in the structure of Ps01740 (128 amino acids), and its side chain forms a salt bridge with Arg-68. The prime indicates that the residue comes from an adjacent monomer.

The crystal structure of Ps01740 inactivated by 2-OP (**Ps01740-2OP**) was determined at a resolution of 2.6 Å via molecular replacement using a monomeric Ps01740 as a search model. Two of the three monomers are covalently modified at Pro-1. In both modified sites, the C-1 carboxylate group of the 2-oxo-3-pentenoate adduct

extends towards the center of the trimer. The overall structures of Ps01740 and Ps01740-2OP are highly homologous, as the RMSD value for the structures is 0.42 Å. However, the conformation of Arg-68 differs in the modified active site of Ps01740-2OP, as the side chain (Nη) does not contact either carboxylate oxygen of Glu-112 as observed in Ps01740. Instead, one carboxylate oxygen of Glu-112 hydrogen bonds with the C-2 oxygen atom (2.9 Å) of the 2-oxo-3-pentenoate adduct on Pro-1 (**Figure 4-2B**). One carboxylate oxygen atom of 2-OP is 3.6 Å from the backbone carbonyl oxygen of Gln-66, but is otherwise free of ordered interactions in the interior of the trimer. On the opposite side of Pro-1, the side chain of His-28 (Nε) is stabilized by the hydroxyl group of Thr-32. The side chain of His-28 (Nε) and Arg-71 (Nη) interact with a water molecule which is 4.5 Å from C-4, opening up the possibility that these these residues might substitute for Glu-112 in water activation for addition to 2-OP.

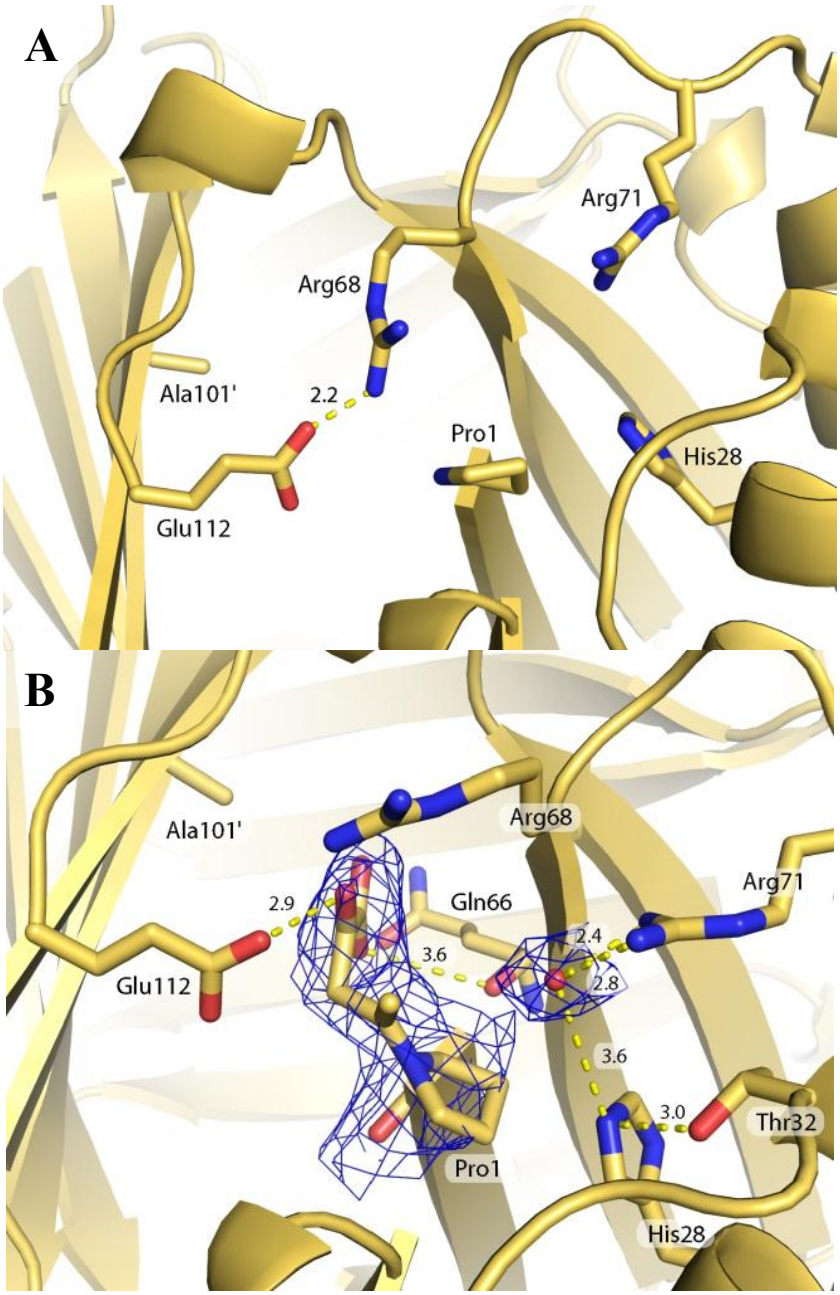


Figure 4-2 Active sites of Ps01740 (2.6 Å resolution) and Ps01740-2OP (2.5 Å resolution).

Distances (yellow dashes) are in Å. **A)** Active site of Ps01740. **B)** $2F_o$-F_c electron density map of 2-OP adduct at Pro-1 and nearby ordered water, contoured at 0.8 σ.

Data Collection	Ps01740	Ps01740-2OP
PDB entry	5UIF	5UMA
space group	*C*2	*C*2
cell dimensions		
a, b, c (Å)	80.7, 58.9, 84.8	81.1, 59.6, 85.1
α, β, γ (deg)	90.0, 110.4, 90.0	90.0, 110.4, 90.0
resolution (Å)	50.00 - 2.58	50.00-2.46
R_{sym} (%) (linear R factor)	0.524 (0.118)	0.484 (0.090)
$I/\sigma I$	7.9 (1.5)	8.2 (1.6)
completeness (%)	96.0 (76.3)	72.9 (65.3)
Refinement		
resolution (Å)	46.481 - 2.574	40.178-2.458
no. of reflections	11402	10226
R_{work}/R_{free} (%)	21.92/27.16	19.17/24.12
no. of atoms		
protein	2520	2567
water	57	36
Average *B* factor (Å^2)		
protein	52.3	47.7
water	44.7	46.8
rmsd		
bond lengths (Å)	0.002	0.002
bond angles (deg)	0.464	0.575
Ramachandran plot (%)		
residues in most favored regions	99.09	98.79
residues in additional allowed regions	0.91	1.21
residues in disallowed regions	0.00	0.00

Data for the last resolution shell are given in parentheses.

Table 4-2 Crystallographic data and refinement statistics for Ps01740 and Ps01740-2OP.

4.3.2 Crystal Structure of Pt0534-2OP

Pt0534 crystallized in the *P*1 space group and contained 6 monomers per asymmetric unit. The structure of Pt0534 inactivated by 2-OP (**Pt0534-2OP**) was obtained at a resolution of 2.0 Å via soaking experiments. Molecular replacement was performed using a 1.2 Å resolution Pt0534 structure that lack a resolved C-terminal hairpin. This incomplete structure was determined using a model for Cg10062. In four of the six monomers of Pt0534-2OP, each of the 123 residues of the protein was resolved (the other two monomers have 122 resolved residues). In Pt0534-2OP, only one of the six monomers is modified.

The tertiary structure of the Pt0534-2OP monomers is identical to that of Ps01740, except a β-hairpin is resolved at the C-terminus, after the second β-α-β unit (**Figure 4-3A**). Additionally, a short α-helix follows the hairpin. The trimer of Pt0534-2OP is homologous to Ps01740 and has a three-fold rotational axis of symmetry (**Figure 4-3B**). The conformation of the active site residues does not differ considerably in the unmodified active sites.

All interactions between the 2-oxo-3-pentenoate adduct and Pt0534 are mediated by water (**Figure 4-3C**). The C-2 carbonyl oxygen interacts with a water molecule stabilized by the backbone carbonyl of Gly-111 and side chain oxygen of Tyr-2. One of the carboxylate oxygen atoms of the 2-oxo-3-pentenoate adducts interacts with two water molecules, where one donates a hydrogen bond to the backbone carbonyl of Ala-113, and

the other is in contact with the side chain of Arg-99′ (Nη). The ε nitrogen of the side chain of Arg-123 is positioned 4.0 Å from the 2-oxo-3-pentenoate carboxylate oxygen.

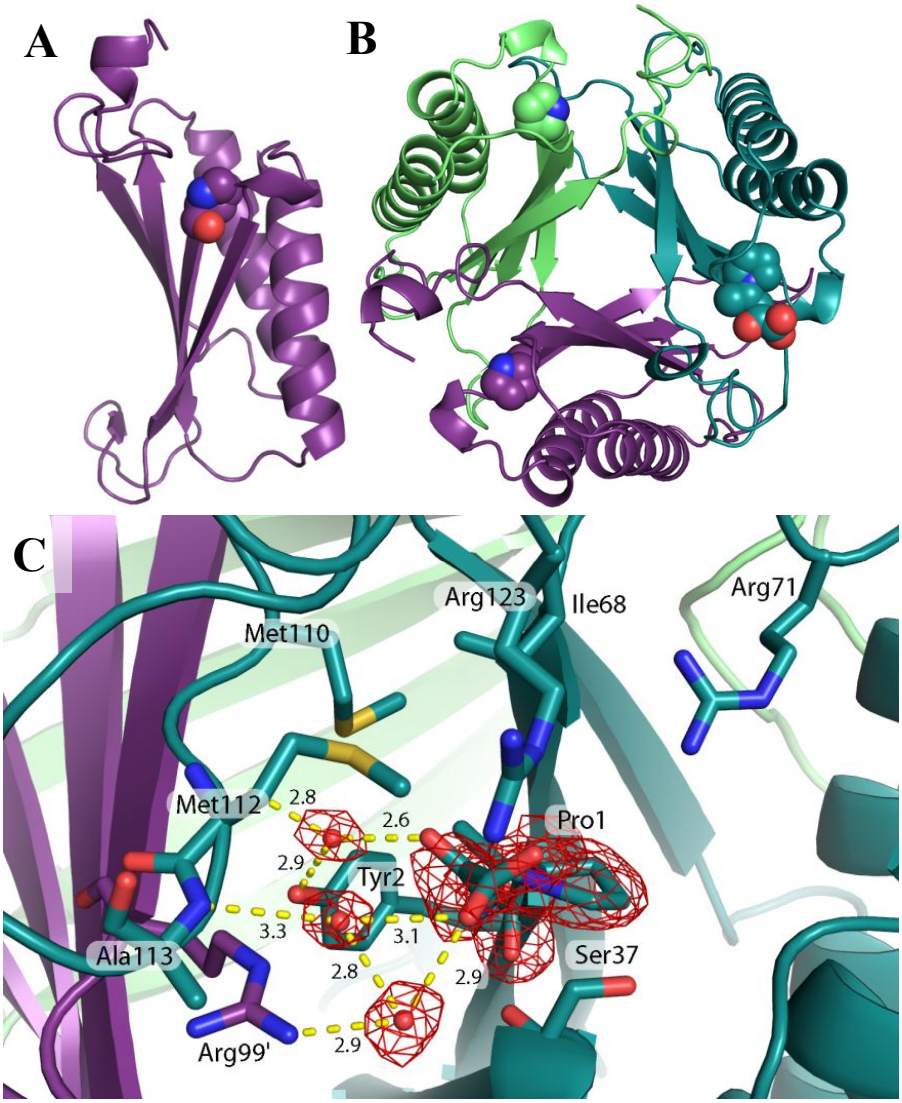


Figure 4-3 The 2.0 Å resolution crystal structure of Pt0534-2OP.

A) Unmodified Pt0534-2OP monomer with Pro-1 atoms shown as spheres. **B)** Pt0534-2OP trimer. Atoms of Pro-1 and 2-OP adduct shown as spheres. **C)** Modified active site. $2F_o$-F_c electron density map surrounding modified Pro-1 and nearby water molecules, contoured at 0.8 σ.

Data Collection	Pt0534-2OP
PDB entry	5UNQ
space group	*P*1
cell dimensions	
a, b, c (Å)	54.214, 57.586, 63.528
α, β, γ (deg)	106.693, 104.539, 94.028
resolution (Å)	50.00 - 1.98
R_{sym} (%) (linear R factor)	0.088 (0.437)
$I/\sigma I$	6.6 (1.6)
completeness (%)	96.0 (85.9)
Refinement	
resolution (Å)	40.830 - 1.976
no. of reflections	47016
R_{work}/R_{free} (%)	20.34/25.18
no. of atoms	
protein	5492
water	645
Average *B* factor (Å^2)	
protein	30.97
water	37.18
rmsd	
bond lengths (Å)	0.003
bond angles (deg)	0.563
Ramachandran plot (%)	
residues in most favored regions	98.89
residues in additional allowed regions	1.11
residues in disallowed regions	0.00

Data for the last resolution shell are given in parentheses.

Table 4-3 Crystallographic data and refinement statistics for Pt0534-2OP.

4.3.3 Crystal Structure of Fused 4-OT

Fused 4-OT crystallizes in the $P2_12_12_1$ space group, and contains 3 monomers per asymmetric unit. The native crystal structure of Fused 4-OT (**F4OT**) was determined at 1.5 Å resolution by molecular replacement using a 1.8 Å resolution Fused 4-OT structure (solved via molecular replacement using Cg10062). Each monomer is composed of two β-α-β motifs, a linker that connects the two units, and a C-terminal motif (**Figure 4-4**). As with Ps01740 and Pt0534, a β-hairpin follows each β-α-β unit in the linker and at the C-terminus, which interacts with neighboring subunits. A 3_{10} helix is positioned between each β-hairpin and the preceding β-α-β unit. All 127 residues in each monomer of Fused 4-OT structure are resolved.

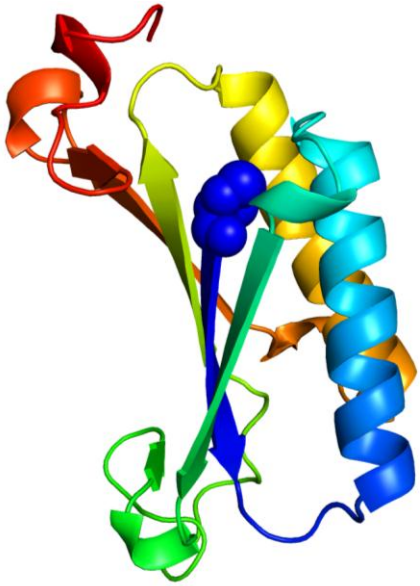


Figure 4-4 Monomer of the 1.5 Å resolution Fused 4-OT (F4OT) crystal structure.

The trimeric structure of Fused 4-OT is unsusual because it does not have three-fold rotational symmetry present in all other known trimeric structures within the TSF. One monomer (Chain A) is "flipped" or inverted relative to the other two monomers in the trimer (**Figure 4-5A**). In this configuration of monomers, each interface in the trimer is unique. One interface has two Pro-1 residues and two Arg-39 residues and is designated **Interface A** (**Figure 4-5B**). The two-fold symmetry in this interface arises from the inversion of one Fused 4-OT monomer. Another interface is asymmetrical and is designated **Interface B**, where Pro-1 is near Arg-104′ of the adjacent monomer (**Figure 4-5C**), replacing Arg-39′ in active sites of Interface A. This interface can be described as "intact," as it would exist between each monomer if inversion did not occur, as is true for other trimeric TSF members. A third interface has two-fold rotational symmetry with no Pro-1 residues, and is designated **Interface C** (**Figure 4-5D**). Again, the prime indicates that the residue comes from an adjacent monomer.

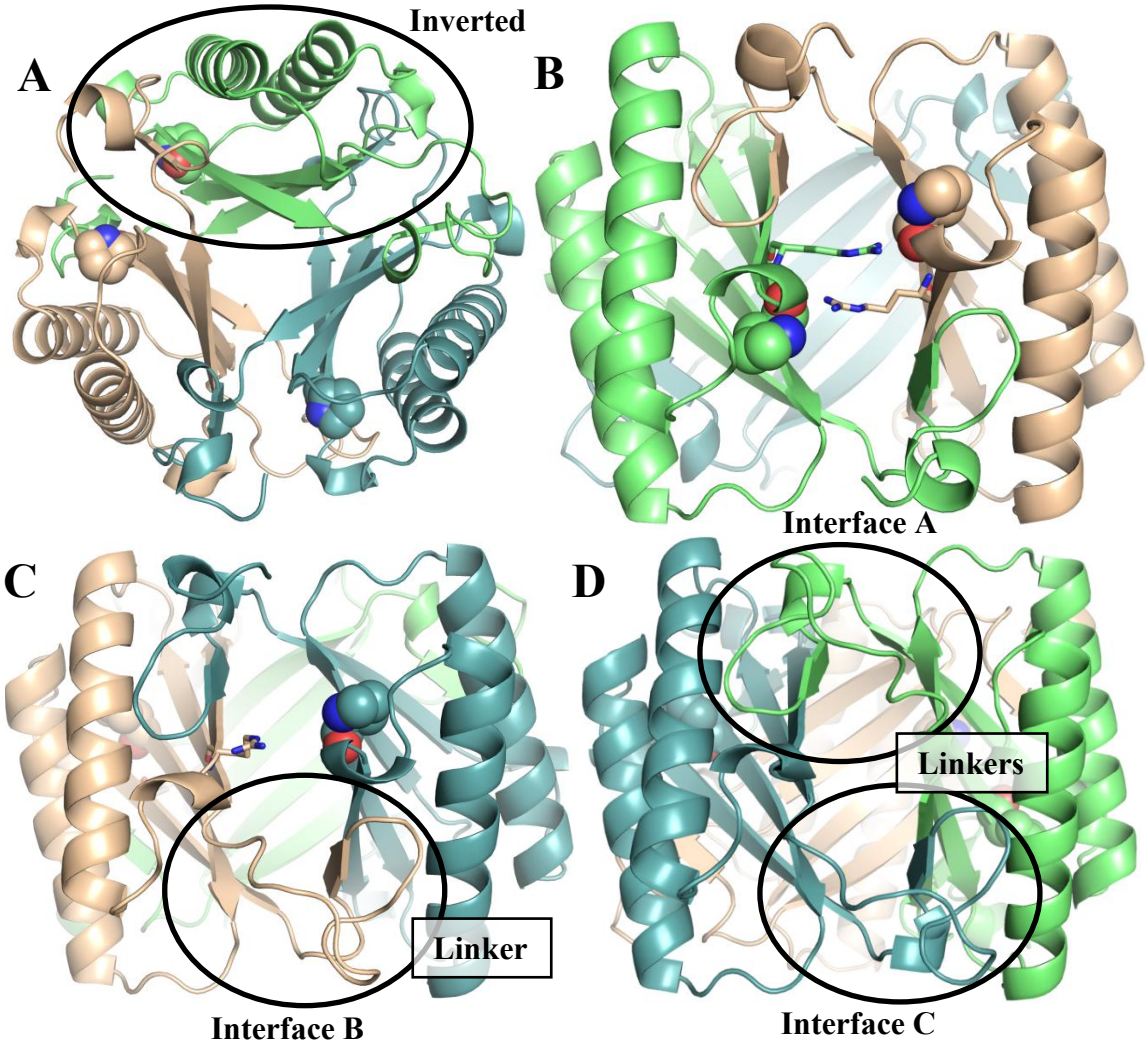


Figure 4-5 The 1.5 Å resolution F4OT crystal structure.

Pro-1 is shown as spheres. **A)** *Face* view of the F4OT hexamer and identification of the inverted monomer. **B)** *Side* view of Interface A. Arg-39 is shown as sticks. **C)** *Side* view Interface B with indicated linker. Arg-104 (sticks) replaces Arg-39 of Interface A. **D)** *Side* view of Interface C with indicated linkers.

The asymmetry observed in the trimer prompted a closer examination of monomers for a better understanding of the causes or consequences of the inverted

configuration. Superposition of the F4OT monomers (**Figure 4-6A**) reveals that Chain C deviates significantly from Chain A (inverted monomer) and Chain B near the β-hairpin region of the linker. The structural deviation observed in the Fused 4-OT monomers is greatest near the beginning of the second β-α-β unit, where a difference in peptide configuration is observed (**Figure 4-6B**). The Leu-65/Pro-66 peptide bond is in the *cis* configuration for both Chain A (lime) and Chain B (wheat), but is *trans* for Chain C (light teal). The difference is accompanied by a 3 Å backbone displacement N-terminal to Pro-66 of the second β-α-β unit and a 180° difference in side chain orientation of Leu-65. Unlike Leu-65 of Chain A or B, the side chain of Leu-65 in Chain C extends towards Ala-98 and Pro-99, where the backbone is displaced up to 2 Å relative to the Chain A and B.

The alternate conformations of the Leu-65/Pro-66 peptide bond may be related to the asymmetrical arrangement of subunits in Fused 4-OT. The linker regions of Chains A and B are near an interface with two-fold rotational symetry (Interface C, **Figure 4-5D**), and their environments are identical, consistent with their common peptide bond configuration at Leu-65/Pro-66 (*cis*). The linker region of Chain C is near Interface B in a different environment than the other linkers (**Figure 4-5C**), consistent with the difference in peptide bond configuration at Leu-65/Pro-66 (*trans*).

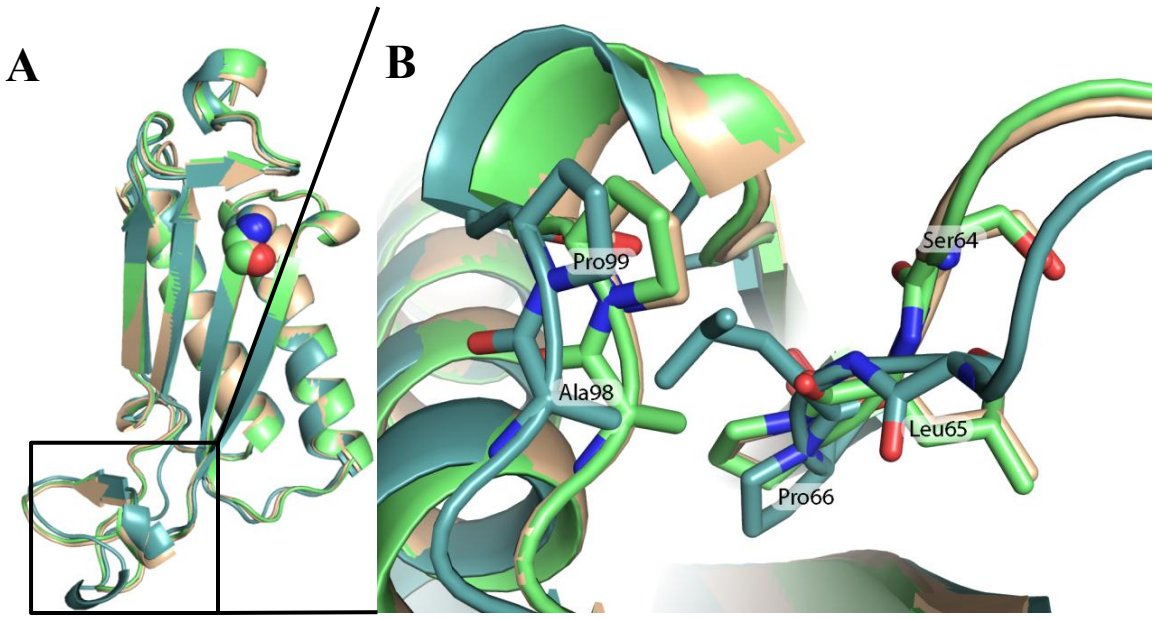


Figure 4-6 Comparison of F4OT monomers.

The structures depicted are the 1.8 Å resolution Fused 4-OT structure solved prior to the 1.5 Å resolution F4OT structure (RMSD 0.19). The Leu-65/Pro-66 peptide bond configurations are the same for each chain in the two structures **A)** Superposition of F4OT monomers. RMSD for Chain A (lime) and Chain B (wheat) is 0.38 Å, the RMSD value for Chain A and Chain C (light teal) is 0.71 Å. **B)** Close-up of superposition with Pro-66 in *cis*-configuration (Chains A and B, lime and wheat, respectfully) and *trans*-configuration (Chain C, light teal).

The conformation of the arginine side chains in each active site of Fused 4-OT are different (**Figure 4-7A**), including Arg-104 and Arg-39, which differ in identity in Interface A when compared to Interface B (**Figure 4-7B**). In Interface A (**Figure 4-5B**), the side chain of Arg-39′ (Nη) hydrogen bonds to the side chain hydroxyl group of Thr-2′ and the backbone carbonyl oxygen of Ser-37 (**Figure 4-7C**).

In the Interface B (**Figure 4-5C**), Val-67′ replaces Thr-2′ present in the Interface A, and a hydrogen bond cannot be made with the side chain of Arg-39 (Nη) (**Figure 4-3D**). Furthermore, the backbone carbonyl oxygens of Ala-102′ and Ser-64′ of the linker region form a hydrogen bond with the side chain of Arg-39 (Nε and Nη, respectively). Arg-104′ does not interact with Thr-2 as observed for Arg-39′ in Interface B (**Figure 4-3B**). Instead, the side chain of Arg-104′ (Nη) interacts with backbone carbonyl groups of Gln-101′ and Thr-103′.

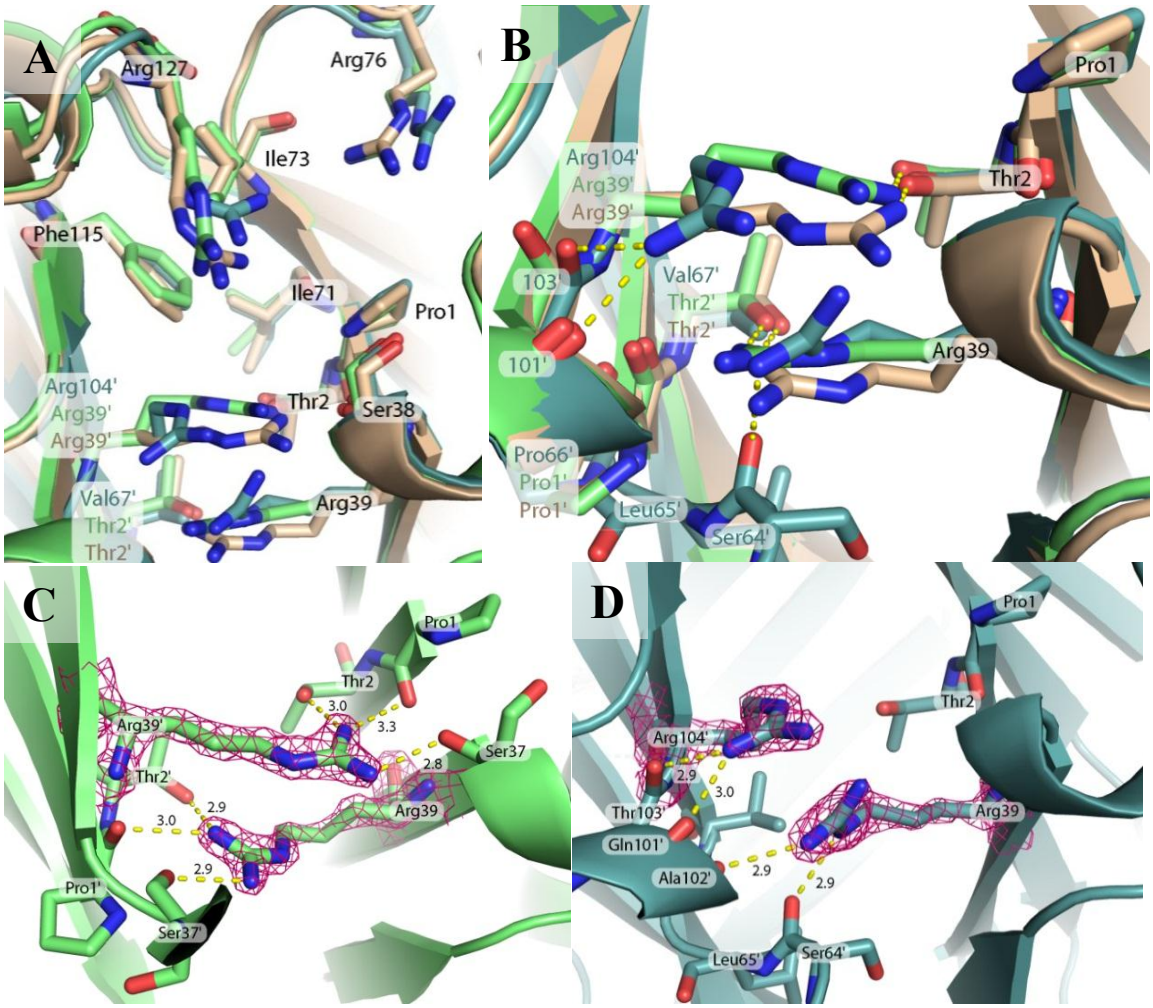


Figure 4-7 Comparison of active sites in F4OT.

Images in **A** and **B** depict a 1.8 Å resolution Fused 4-OT structure determined prior to the solution of the 1.5 Å resolution F4OT structure (RMSD 0.19). Hydrogen bonding interactions in **A** and **B** are the same for both structures. **A)** Superposition of active sites of Interface A (lime and wheat) and Interface B (deep teal). **B)** Close-up of superposition showing Arg-39 and Arg-104 conformations with surrounding residues. **C)** $2F_o$-F_c electron density map (1.5 Å resolution F4OT structure) surrounding Arg-39 residues in Interface A, contoured at 2.0 σ. **D**) $2F_o$-F_c electron density map of Arg-39 and Arg-104′, contoured at 2.0 σ in the Interface B. Side chains of Thr-103′, Gln-101′, and Ala-102′ are oriented away from Arg-39 and Arg-104′ and are not shown.

Data Collection	F4-OT
PDB code	5UN4
space group	$P2_12_12_1$
cell dimensions	
a, b, c (Å)	47.8, 67.2, 94.8
α, β, γ (deg)	90.0, 90.0, 90.0
resolution (Å)	50.00 - 1.49
R_{sym} (%) (linear R factor)	0.083 (0.698)
$I/\sigma I$	25.5 (1.8)
completeness (%)	99.7 (99.7)
Refinement	
resolution (Å)	42.695 - 1.488
no. of reflections	50550
R_{work}/R_{free} (%)	19.531/21.01
no. of atoms	
protein	2751
water	329
Average *B* factor (Å^2)	
protein	25.9
water	37.04
rmsd	
bond lengths (Å)	0.007
bond angles (deg)	1.113
Ramachandran plot (%)	
residues in most favored regions	99.47
residues in additional allowed regions	0.53
residues in disallowed regions	0.00

Data for the last resolution shell are given in parentheses.

Table 4-4 Crystallographic data and refinement statistics for F4OT.

4.3.4 Crystal Structures of F4OT-2HM, F4OT-2OP, and F4OT-PP

A crystal structure of Fused 4-OT in complex with 2-HM was solved to 1.9 Å resolution (**F4OT-2HM**). This is the first crystal structure of any TSF member bound to 2-HM. The ligands bound may be the products 2-keto-3-hexenedioate or 2-keto-4-hexenedioate, or a mixture of substrate and products, but these compounds are tautomers, and their identity cannot be determined based on electron density maps. The ligand is bound with partial occupancy in near one Pro-1 in Interface A (**Figure 4-8B,C**) and the Pro-1 in interface B (**Figure 4-8D,E**) and in the linker region of Chain A (**Figure 4-8F**).

In both active sites bound with 2-HM, the C-1 carboxylate of 2-HM hydrogen bonds with the side chain of Arg-76 (Nη) as well as the backbone amide of Ile-73 (**Figure 4-5C,E**). In both sites, C-2 oxygen hydrogen bonds with Pro-1 and a water molecule stabilized by the side chain of Thr-2. Side chains of Ile-73, Phe-115, and Ile-117 are near hydrophobic atoms C-3, C-4, and C-5.

In Interface A, the side chain of Arg-39′ ($N\eta_1$) contacts the 2-HM C-6 carboxylate, and Arg-39′ ($N\eta_2$) is in contact with the side chain of Thr-2 (**Figure 4-5B**). In Interface B, the side chains of Arg-104′ (Nη) and Arg-39 (Nη) hydrogen bond with the C-6 carboxylate, and neither residue interacts with the side chain hydroxyl group of Thr-2 (**Figure 4-5D,E**).

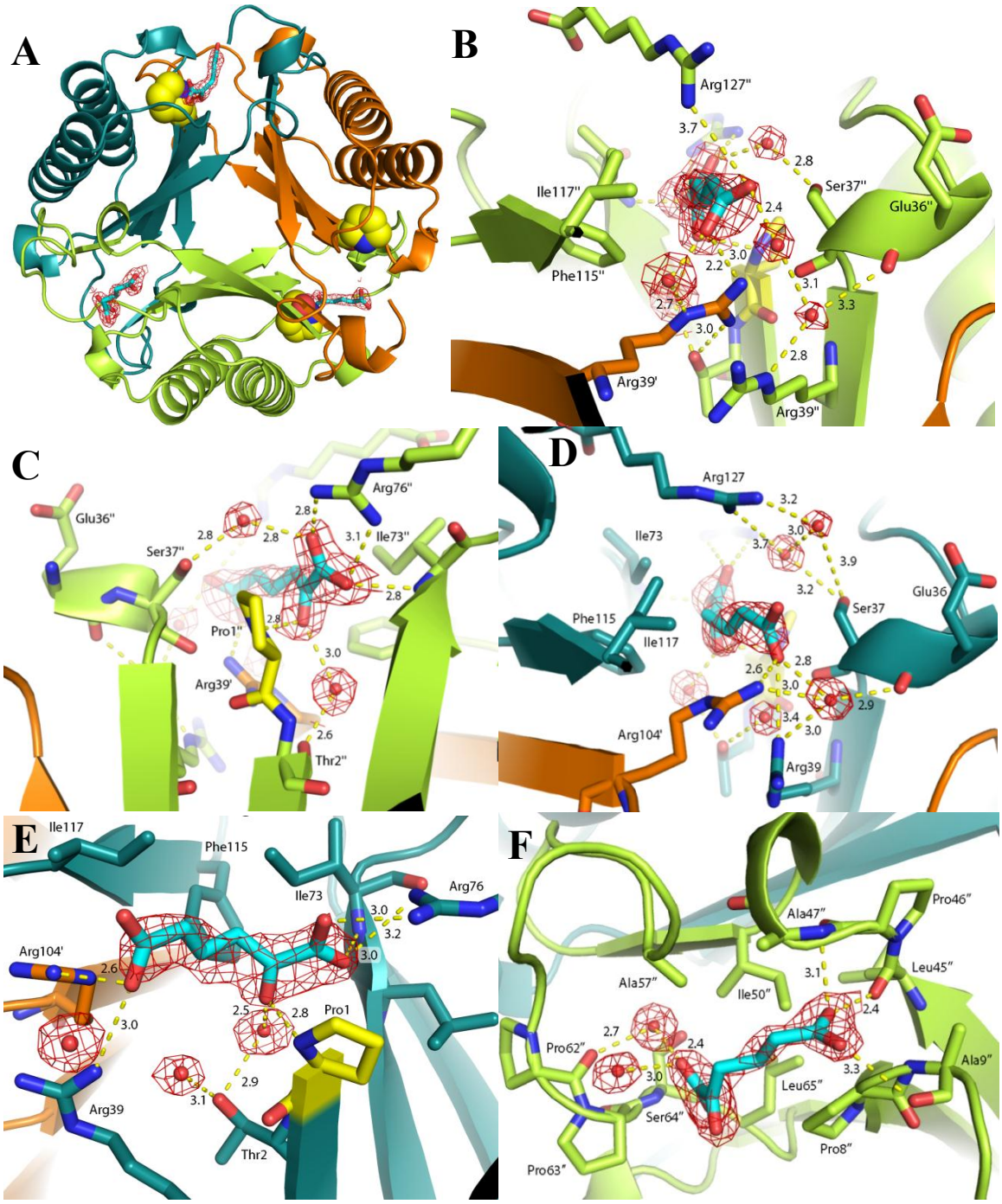


Figure 4-8 The 1.9 Å resolution structure of F4OT-2HM shown with $2F_0$-F_c map contoured at 0.5 σ.

Figure 4-8 (*continued*) Distances (yellow) in Å. **A)** *Face* view of F4OT-2HM. Pro-1 is (yellow) spheres in each monomer. **B,C)** 2-HM (cyan) bound at Interface A. **D,E)** 2-HM bound at Interface B. **F)** 2-HM bound near linker region of Chain A.

Crystals of Fused 4-OT were modified by 2-OP via soaking experiments for the solution of **F4OT-2OP**. The structure was solved at 2.3 Å resolution by molecular replacement using a 1.8 Å native Fused 4-OT structure solved prior to F4OT. All three active sites were modified (**Figure 4-9A**), and the binding modes of the 2-oxo-3-pentenoate adduct were different in each case.

In the Interface B, the carboxylate oxygen atoms of the adduct are interacting with the side chain of Arg-104′ (Nη and Nε) and a water molecule stabilized by Arg-39 (Nη) (**Figure 4-9B**). The C-2 oxygen interacts with the side chain of Arg-127 (Nη).

In the Interface A, the C-2 oxygen atoms of both 2-oxo-3-pentenoate adducts interact with Arg-39′ (Nη and Nε, **Figure 4-9C**). The C-1 carboxylate is oriented differently in the two active sites — the side chain of Arg-127 (Nη) in Chain C interacts with the C-1 carboxylate of the 2-oxo-3-pentenoate adduct whereas the side chain of Arg-76 (Nη) in Chain A interacts with the C-1 carboxylate via a water molecule (**Figure 4-9C**).

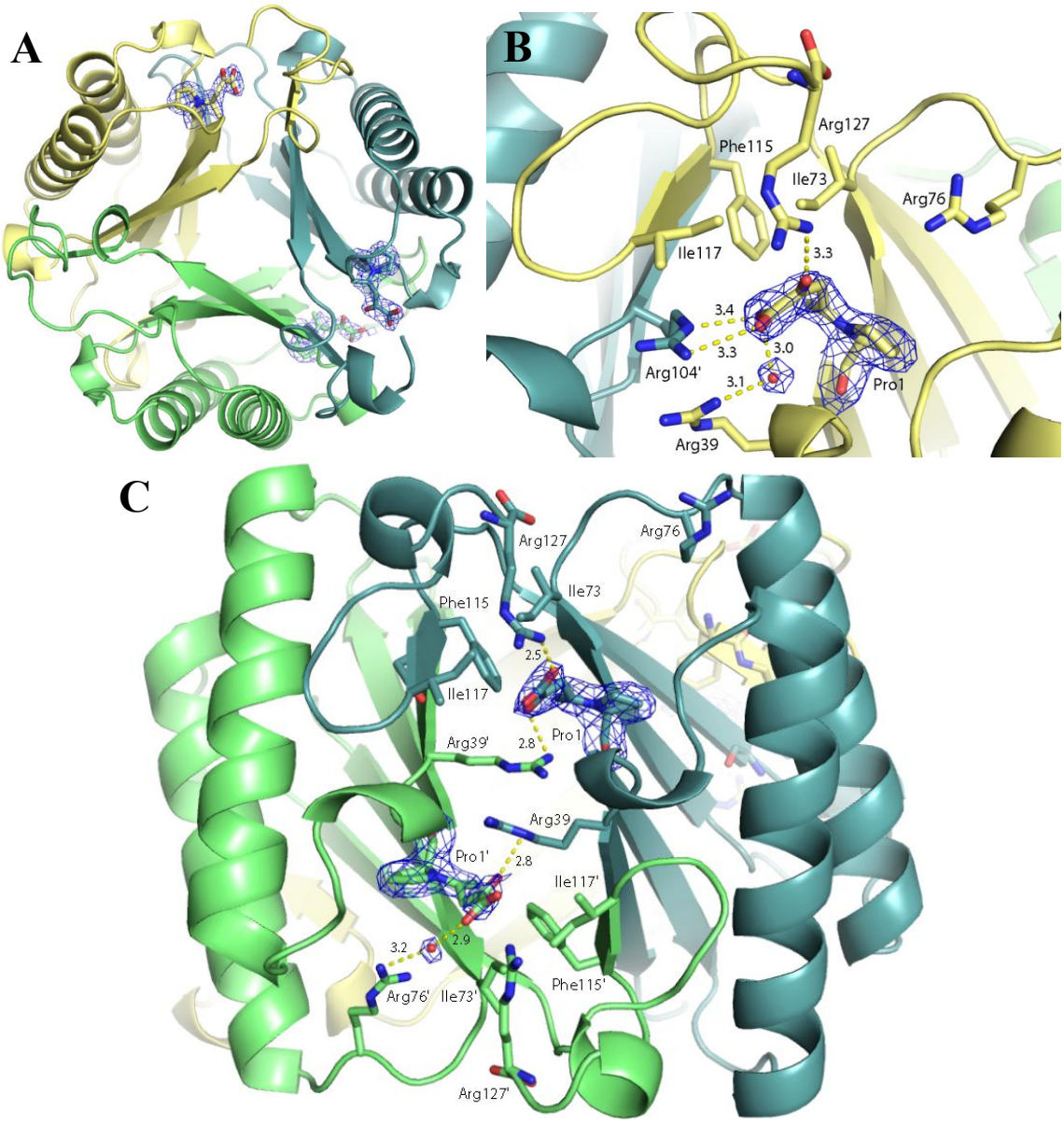


Figure 4-9 The 2.3 Å resolution F4OT-2OP crystal structure.

Continuous $2F_o$-F_c electron density map surrounds the 2-oxo-3-pentenoate and Pro-1, contoured at 1.0 σ. **A)** *Face* view of F4OT-2OP. **B)** Modified active site in Interface B. **C)** Interface A with two modified Pro-1 residues.

Differences in conformation of ligands and arginine residues are observed when active sites of F4OT-2OP and F4OT-2HM are superimposed. In the superposition of F4OT-2HM active sites, Arg-76 of Chain B in Interface B does not extend towards Pro-1 like the other chains, nor is 2-HM bound in this active site (**Figure 4-10A**). The differences in binding of the C-6 carboxylate of 2-HM are discussed above. In the alignment of F4OT-2OP active sites, the alternative positioning of Arg-39′ and Arg-39 in Interface A relative to Arg-104′ and Arg-39 in Interface B is accompanied by an alternate orientation of the 2-oxo-3-pentenoate (**Figure 4-10B**) where the carboxylate adduct (yellow orange) hydrogen bonds with the side chain of Arg-104′ (teal).

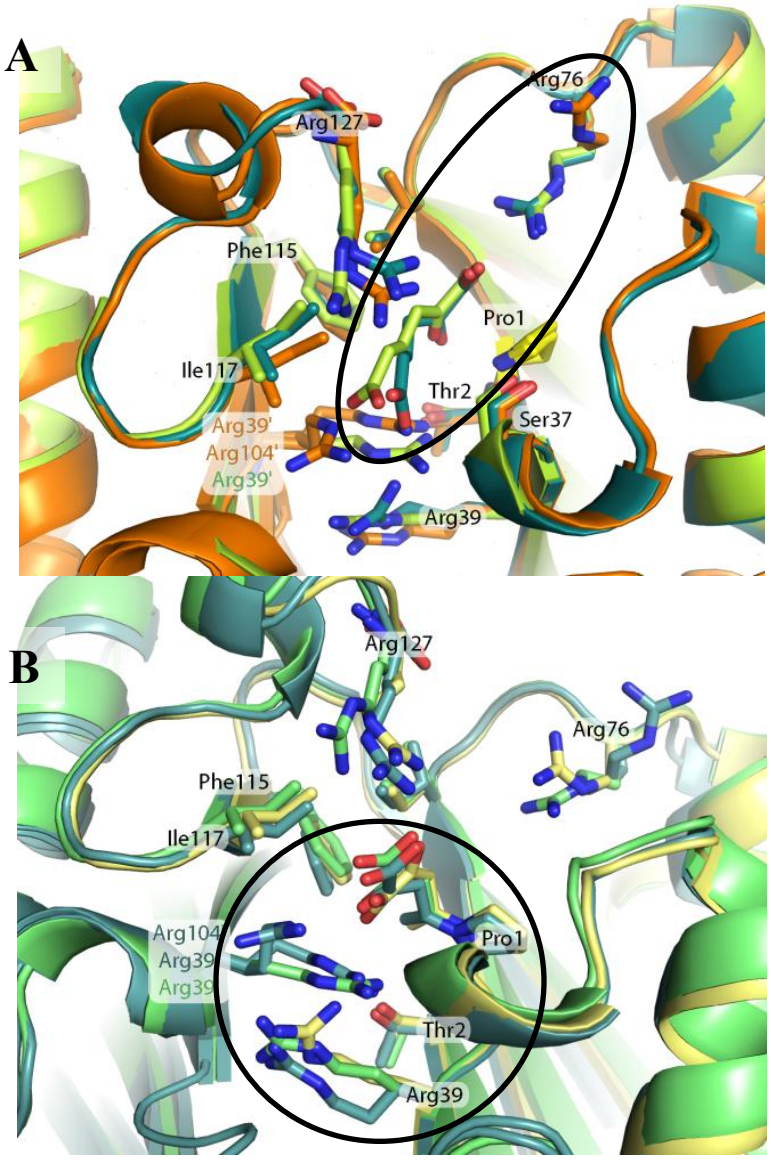


Figure 4-10 Superposition of **A)** F4OT-2HM and **B)** F4OT-2OP active sites.

A) The alternate conformation of Arg-76 in Chain B (orange) in Interface A is concomminant with the lack of bound 2-HM. **B)** The alternate conformations of Arg-39 and Arg-104′ side chains in Interface B (yellow orange, and teal, respectively) relative to the Arg-39 and Arg-39′ side chains in Interface A (green and teal) is accompanied by alternate conformations of the 2-oxo-3-pentenoate adduct (yellow orange at Interface B, and green/teal at Interface A).

The Fused 4-OT structure in complex with phenylpyruvate (**F4OT-PP**) was solved to 2.7 Å resolution using phenylpyuvate-soaked crystals. The $2F_o$-F_c electron density map near Pro-1 in Chain A (at Interface A) indicated phenylpyruvate binds in two positions. Two phenylpyruvate ligands were modeled in the structure with partial occupancy (**Figure 4-11**). The pyruvoyl moiety of one phenylpyruvate ligand is oriented away from Pro-1 and its C-2 carbonyl interacts with a water molecule stabilized by the side chain of Arg-39′ (Nε and $N\zeta_1$). The pyruvoyl moiety of the other phenylpyruvate ligand is oriented towards Pro-1, where the C-1 carboxylate of interacts with the side chain of Arg-39′ ($N\zeta_2$), the prolyl nitrogen of Pro-1, and the backbone carbonyl of Ile-73. The C-2 oxygen atom is interacting with the backbone amide of Ile-73 and the side chain of Arg-76 (Nη). The presence of these interactions and the fact that the prolyl nitrogen of Pro-1 is 3.2 Å from C-3 of phenylpyruvate indicates that the orientation might resemble the catalytic conformation.

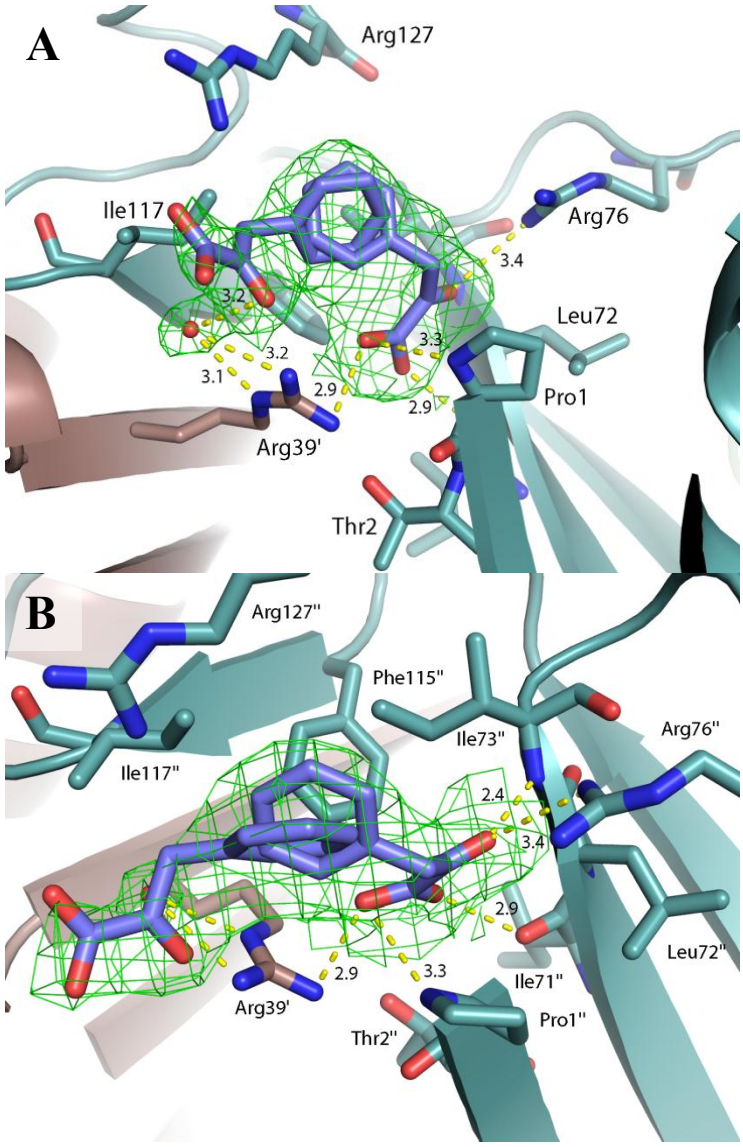


Figure 4-11 Active site of the 2.7 Å resolution F4OT-PP crystal structure.

A, B) Two views of the active site in two conformations in Interface A with $2F_o$-F_c electron density surrounding phenylpyruvate (slate) and a nearby ordered water molecule, contoured at 0.5 σ.

Data Collection	F4OT-2HM	F4OT-2OP	F4OT-PP
PDB code	N/A	N/A	N/A
space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
cell dimensions			
a, b, c (Å)	48.1, 68.6, 96.4	48.1, 69.4, 96.4	48.2, 69.5, 96.1
α, β, γ (deg)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
resolution (Å)	50.00 - 1.88	50.00 - 2.30	50.00 - 2.69
R_{sym} (%) (linear R factor)	0.092 (0.123)	0.271 (0.949)	0.236 (0.738)
$I/\sigma I$	12 (1.6)	6.6 (1.5)	3.153 (1.579)
completeness (%)	95.9 (85.4)	99.9 (100.0)	93.4 (95.7)
Refinement			
resolution (Å)	43.026 - 1.872	36.567 - 2.319	48.067 - 2.663
no. of reflections	25795	14540	8887
R_{work}/R_{free} (%)	17.63/21.82	19.33/23.25	22.22/27.71
no. of atoms			
protein	2783	2780	2751
water	308	190	53
Average *B* factor (Å^2)			
protein	32.8	28.9	42.5
water	40.4	32.5	39.2
rmsd			
bond lengths (Å)	0.004	0.003	0.003
bond angles (deg)	0.771	0.765	0.715
Ramachandran plot (%)			
residues in most favored regions	99.20	98.93	99.20
residues in additional allowed regions	0.80	1.07	0.80
residues in disallowed regions	0.00	0.00	0.00

Data for the last resolution shell are given in parentheses.

Table 4-5 Crystallographic data and refinement statistics for F4OT-2HM, F4OT-2OP, and F4OT-PP.

4.4 DISCUSSION

Ps01740, Pt0534 and Fused 4-OT are linked by sequence identity to the title enzyme in the 4-OT and *cis*-CaaD subgroups. The long monomers of *cis*-CaaD and these linker enzymes may have arisen as a result of duplication and fusion of an ancestral 4-OT-like gene sequence. Crystal structures of the linker enzymes show divergence in active site composition from their better understood homologues 4-OT and *cis*-CaaD, as was expected from sequence analysis (**Figure 4-12**). Arginine residues that bind dicarboxylate substrates such as 2-HM are on opposite sides of Pro-1 in 4-OT (Arg-11′ and Arg-39″), and are substituted for two arginine residues on one side of Pro-1 (Arg-70 and Arg-73) in *cis*-CaaD. His-28, Glu-114, and Tyr-103′ are other key catalytic residues in *cis*-CaaD. The structure-function analysis of the contemporary linkers with intermediate functional properties of both *cis*-CaaD and 4-OT might reveal insight the changes in architecture and active site composition during the proposed evolution from the 4-OT-like enzymes that act on diacid substrates to *cis*-CaaD-like enzymes which process monoacid compounds.

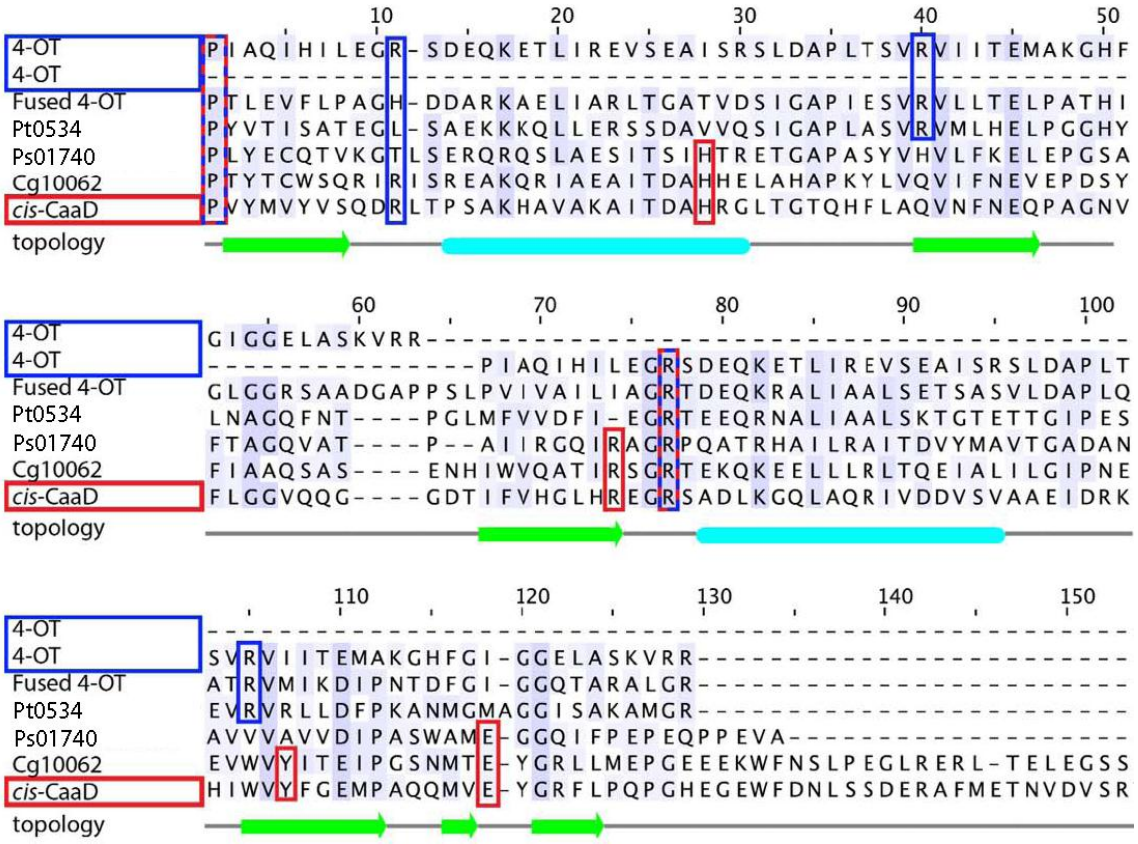


Figure 4-12 Sequence alignment of enzymes of the 4-OT and *cis*-CaaD subgroups.

Active site residues for 4-OT are boxed in blue and active site residues for *cis*-CaaD are boxed in red, adapted from (Davidson *et al.*, *In Preparation*).

4.4.1 Ps01740 Structure and Function

Ps01740 has no activity with *cis*-3-chloroacrylate (LeVieux, *et al., In Preparation*), despite its close sequence similarity with *cis*-CaaD. The key residues in *cis*-CaaD (Pro-1, Arg-70, Arg-73, Glu-114, and His-28) are conserved in Ps01740

(equivalent residues are Pro-1, Arg-68, Arg-71, Glu-112, and His-28, **Figure 4-13B**), and the monomeric RMSD for *cis*-CaaD and Ps01740 is 1.11 Å (**Figure 4-13A**). However, Tyr-103′ in *cis*-CaaD is not conserved in Ps01740, which may explain its lack of *cis*-CaaD activity as the residue is believed to assist Glu-114 (*cis*-CaaD numbering) in the activation of water (Poelarends *et al.*, 2003). The importance of Tyr-103′ is further substantiated by studies of Cg10062, an enzyme that has all 6 catalytic residues found in *cis*-CaaD (34% sequence identity), including Tyr-103′. Cg10062 exhibits *cis*-CaaD activity (k_{cat}/K_M of 184 $M^{-1}s^{-1}$), albeit two orders of magnitute less than k_{cat}/K_M for *cis*-CaaD (Poelarends *et al.*, 2008).

Ps01740 does not process 2-hydroxymuconate (2-HM), likely because it lacks Arg-39 (4-OT numbering) to stabilize the second carboxylate of the diacid. However, the enzyme catalyzes the tautomerization of phenylpyruvate at a rate of 3.5×10^5 $M^{-1}s^{-1}$ (**Table 4-6**). Ps01740 has Pro-1 and Arg-71, which are equivalent to catalytic residues Pro-1 and Arg-11′ in 4-OT (**Figure 4-13C**), and the RMSD for the 4-OT dimer and Ps01740 is 1.23 Å (**Figure 4-13A**). The active site residues Arg-39″ and Phe-50′ in 4-OT are not conserved in Ps01740, and are replaced by Val-99′ and Ala-101′.

Ps01740 does not have a proline residue at the N-terminus of the second β-α-β unit which should align with Pro-1 of 4-OT (**Figure 4-13C**), as observed for Fused 4-OT (**Figure 4-17C**). However, Pro-60 of Ps01740 is located in the linker region two residues N-terminal to this position.

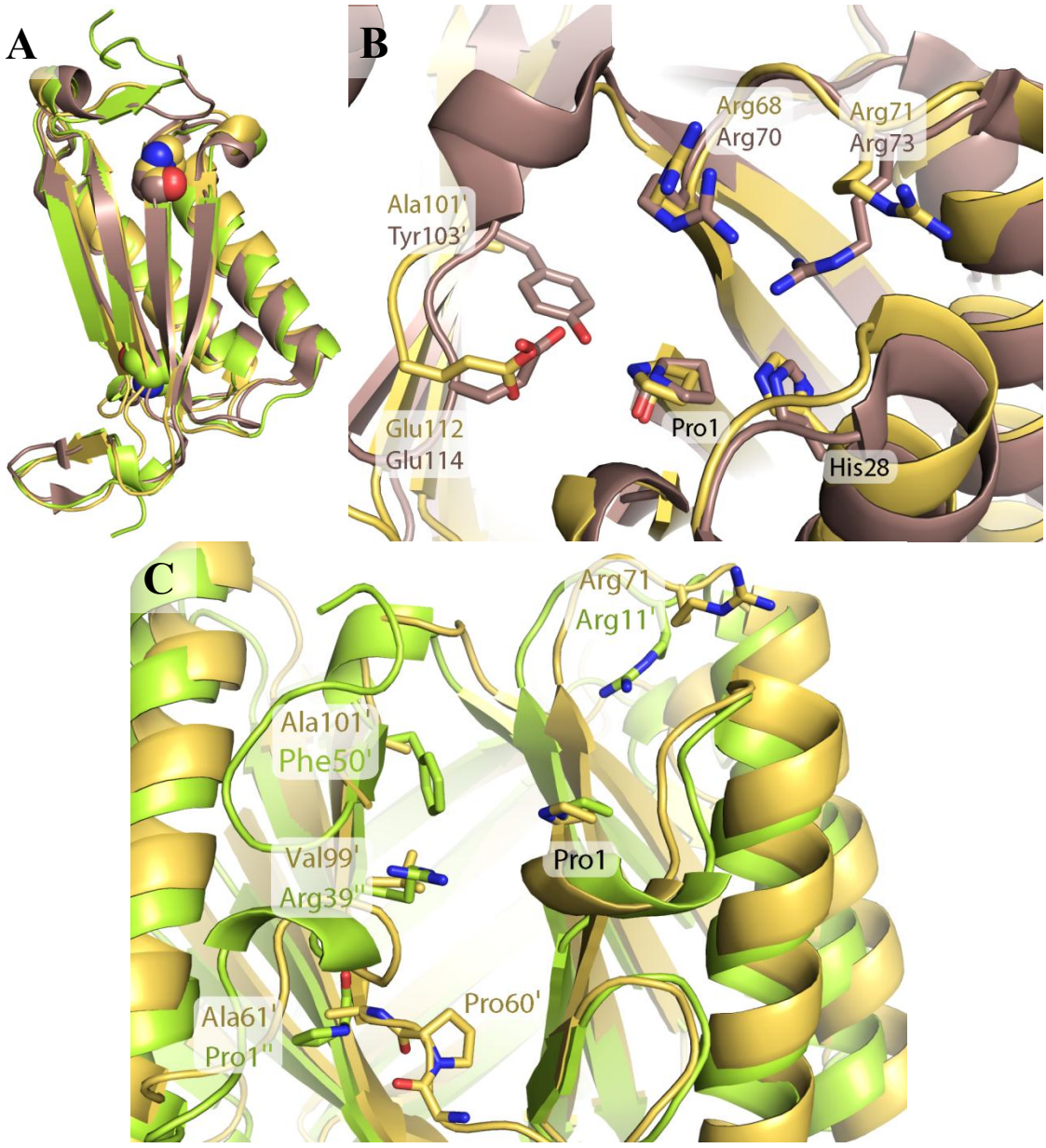
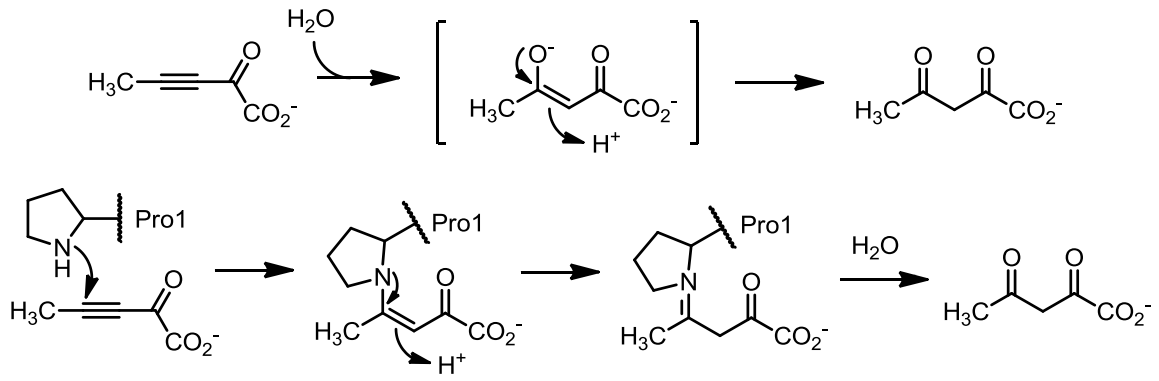


Figure 4-13 Superposition of Ps01740, *cis*-CaaD, and 4-OT.

A) Structural alignment of *cis*-CaaD (dark salmon, PDB entry 3MF8) monomer and the 4-OT dimer (limon, PDB entry 1OTF) to Ps01740 monomer (yellow orange). **B)** Active site comparison of Ps01740 (yellow orange) and *cis*-CaaD (dark salmon). **C)** Active site comparison of Ps01740 (yellow orange) and 4-OT (limon).

Covalent modification of the prolyl nitrogen of Pro-1 in Ps01740 by 2-OP was confirmed by MS (LeVieux, *et al.*, *In Preparation*) and by the solution of a crystal structure of Ps01740-2OP. Comparison of Ps01740-2OP and structures of *cis*-CaaD and Cg10062 modified by 2-oxirane-2-carboxylate (**Figure 4-14A**) reveal an alternative positioning for the 2-oxo-3-pentenoate adduct in the active site of Ps01740 relative to 2-hydroxypropanoate adducts of *cis*-CaaD and Cg10062, where the carboxylate is buried near the center of the trimer. The "in" conformation of the 2-oxo-3-pentenoate may related to the absence of Tyr-103 which gives space to accommodate the 2-oxo-3-pentenoate adduct.

Ps01740 is a member of the *cis*-CaaD subgroup, where the studied members have higher pK_a values (~9.2 for *cis*-CaaD) for Pro-1, and can hydrate 2-OP. Since both hydration and inactivation occur for Ps01740, Pro-1 might have an intermediate pK_a value (~7-8) as was postulated for the *cis*-CaaD homologue known as MsCCH2 (Baas, *et al.*, 2011). The altered binding mode of the 2-oxo-3-pentenoate in Ps01740, as observed in the Ps01740-2OP structure, might be related to the low hydratase activity. Arg-71 and His-28 interact with a water molecule near the modified Pro-1 (**Figure 4-2B**), and are alternate candidates to Glu-112 for water activation. The 2-oxo-3-pentenoate adduct formed at Pro-1 in the crystal structures of Ps01740-2OP and BP4401-2OP (**Section 5.3**), both enzymes which hydrate 2-OP (and are modified by the compound), opens the possibility for covalent catalysis in these enzymes (**Scheme 4-1B**).



Scheme 4-1 A) Direct and **B)** covalent hydration mechanism of Ps01740 using 2-OP.

In the 4-OT co-crystal structure covalently modified by 2-OP, the side chain Arg-39′ (Nε and Nη) hydrogen bond to C-2 oxygen and carboxylate oxygen, respectively (**Figure 4-14B**). Ps01740 does not have an equivalent arginine residue to stabilize the carboxylate of the 2-oxo-3-pentenoate adduct in Ps01740-2OP. The lack of the arginine and presence of Tyr-38 nearby forces the 2-oxopentenoate adduct to the "in" conformation where the adduct extends towards the center of the trimer in a water-filled pocket that exists in the absence of a Tyr-103′ (*cis*-CaaD numbering).

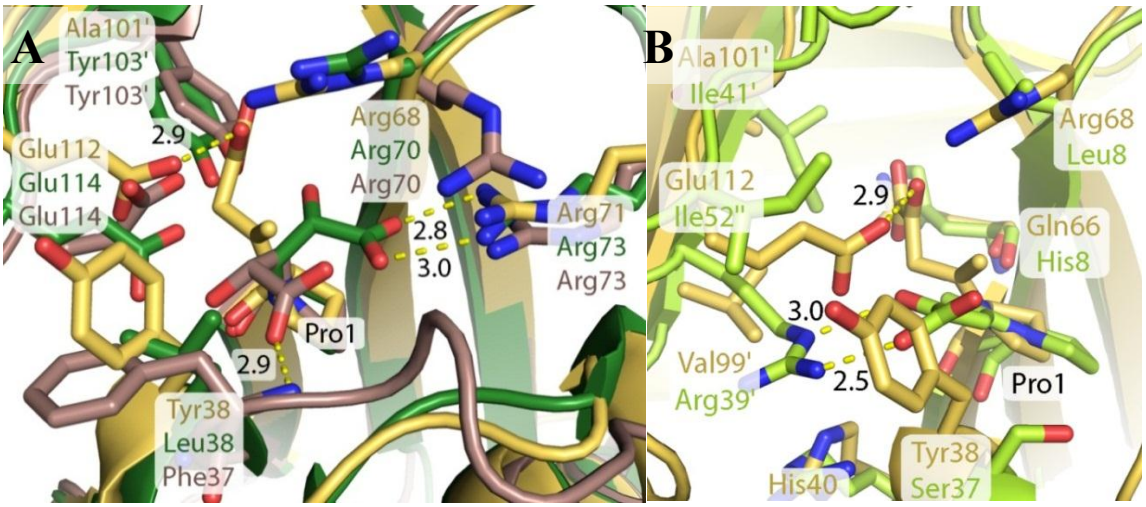


Figure 4-14 Superposition of Ps01740-2OP and inactivated homologues.

A) Overlay of Ps01740-2OP (yellow orange), *cis*-CaaD-2-hydroxypropanate (dark salmon, PDB entry 3MF7) and Cg10062-2-hydroxypropanate (forest, PDB entry 3N4H) active sites. **B)** Active site superposition of Ps01740-2OP (yellow orange) and 4-OT inactivated by 2-OP (limon, PDB entry 1BJP).

4.4.2 Pt0534 Structure and Function

Pt0534 is a member of the 4-OT subgroup, and has trace levels of *cis*-CaaD activity (monitored over several weeks). Unlike Ps01740, the identity of active site residues of *cis*-CaaD and Pt0534 is low, as only Pro-1 and Arg-71 in Pt0534 are conserved (equivalent to Pro-1 and Arg-73 in *cis*-CaaD, **Figure 4-15B**). In addition, the tertiary structure of Pt0534-2OP deviates more from *cis*-CaaD (monomer RMSD: 1.39 Å) than for 4-OT (Pt0534-2OP monomer and 4-OT dimer RMSD: 0.78 Å, **Figure 4-15A**).

Pt0534 is inactivated by 2-OP and has no hydration activity with the acetylene compound, consistent with other members of the 4-OT subgroup.

Pt0534 is an efficient tautomerase, as the k_{cat}/K_M values for 2-HM and phenylenolpyruvate (3.5×10^7 and 1.3×10^6 $M^{-1}s^{-1}$, respectively, Baas *et al.*, *In Preparation*) are comparable for that of 4-OT (5.5×10^7 $M^{-1}s^{-1}$ and 3.7×10^5 $M^{-1}s^{-1}$, respectively, Chen *et al.*, 1992, Burks *et al.*, 2010). The key residues conserved in Pt0534 are Pro-1, Arg-71, and Arg-99′, are equivalent to catalytic residues Pro-1, Arg-11′, Arg-39″, respectively, in 4-OT (**Figure 4-15C**). The active site residue Met-110 in Pt0534 replaces Phe-50′, indicating the residue can be exchanged for another bulky amino acid, as evident in the 4-OT homologue TomN, where Phe-50′ is replaced by tryptophan (Burks, *et al.*, 2011). Pro1″ in 4-OT is not conserved in Pt0534, but Pro-59 is only two residues N-terminal to the start of the second β-α-β unit. The equivalent residue for Arg-61′ in 4-OT is present in Pt0534 (the C-terminal residue Arg-123), and is resolved in the native structure.

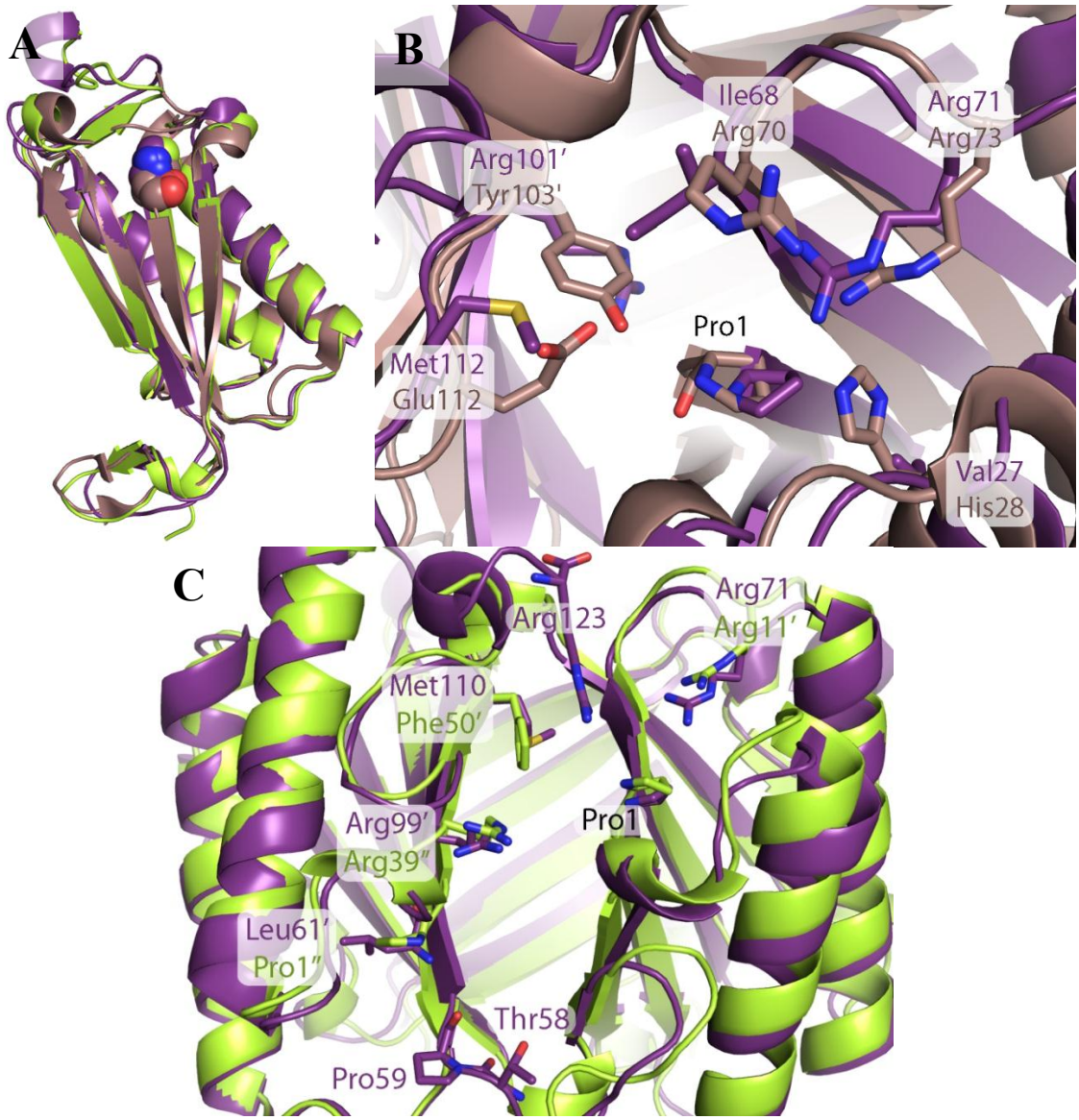


Figure 4-15 Superposition of Pt0534-2OP with 4-OT and *cis*-CaaD.

A) Structural alignment of Pt0534-2OP monomer (unmodified, violet purple) to 4-OT dimer (limon, PDB entry 1OTF) and *cis*CaaD monomer (dark salmon, PDB entry 3MF8), with Pro-1 residues shown as spheres. **B)** Comparison of Pt0534-2OP (unmodified site) and *cis*-CaaD active sites. **C)** Comparison of Pt0534-2OP (unmodified site) and 4-OT active sites.

The 2-oxo-3-pentenoate adduct bound in the modified active site of Pt0534-2OP does not adopt the "in" conformation observed for Ps01740-2OP (**Figure 4-16**). Instead, the adduct is positioned similarly to the ligand in the structure of 4-OT structure inactivated by 2-OP, where the C-1 carboxylate group of the 2-oxo-3-pentenoate adduct extends towards the side chain of Arg-123 (Nε and Nη) (**Figure 4-16**). In 4-OT, the C-1 carboxylate of the 2-oxo-3-pentenoate interacts with the side chain of Arg-61′ (Nε and Nη). The distance is greater in Pt0534-2OP, as the closest atomic distance between the two groups is 4.0 Å. The 2-oxo-3-pentenoate adduct makes direct contact to Arg-39″ in 4-OT, whereas all interactions between the adduct and Pt0534 are mediated by water. Although the interaction of the C-1 carboxylate of the 2-oxo-3-pentenoate adduct and Arg-123 in Pt0534 is less direct, and 4-OT does not require the equivalent residue for activity (Harris, *et al.,* 1999), this C-terminal residue may participate in 2-HM tautomerization in Pt0534, as was observed for TomN (R61A mutant showed a 5-fold decrease in k_{cat}, Burks *et al.*, 2011), and Fused 4-OT (100-fold decrease in k_{cat}/K_M, **Section 4.4.4**).

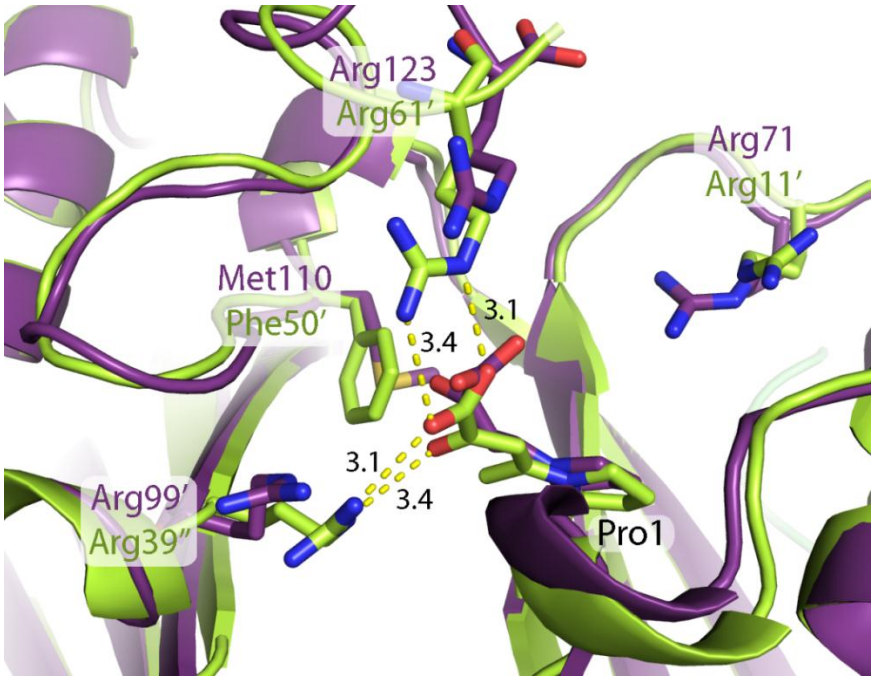


Figure 4-16 Alignment of Pt0534 (deep teal and violet purple) and 4-OT labeled with 2-oxo-3-penteonate adduct (orange, PDB entry 1BJP).

4.4.3 Fused 4-OT Structure and Function

The tertiary structure of the native Fused 4-OT (**F4OT**) monomer resembles the 4-OT dimer more than it does the *cis*-CaaD monomer (RMSD of F4OT and 4-OT dimer is 0.65 Å and RMSD for F4OT and *cis*-CaaD is 5.2 Å, **Figure 4-17A**). One monomer of F4OT is inverted relative to the others, which could result from the different peptide bond configurations at Leu-65/Pro-66, creating two equivalent active sites in the Interface A (**Figure 4-5B**) and one active site in Interface B (**Figure 4-5C**) that is not equivalent to the other two. Though sequence identity of Fused 4-OT and *cis*-CaaD is lowest out of the

three enzymes studied, Fused 4-OT has trace *cis*-CaaD activity (monitored over several weeks, Levieux J, *et al.*, *In Preparation*). The catalytic residues Pro-1 and Arg-76 in *cis*-CaaD are conserved in Fused 4-OT (Pro-1 and Arg-73, respectively, **Figure 4-17B**).

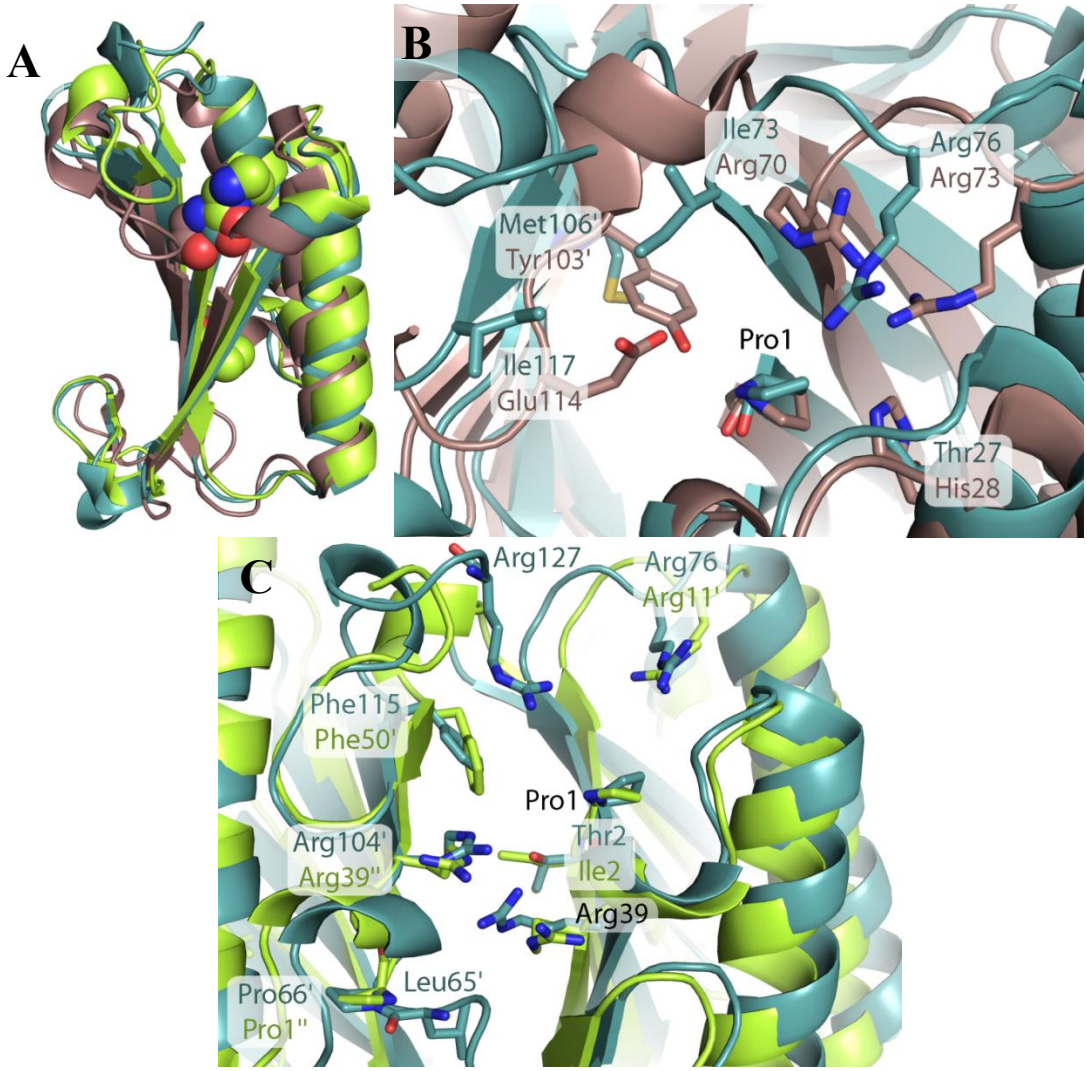


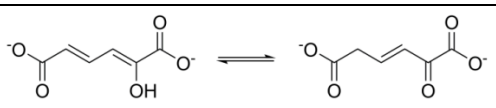
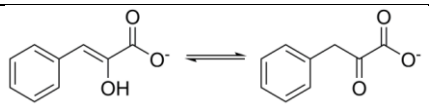
Figure 4-17 Superposition of F4OT with 4-OT and *cis*-CaaD.

A) Structural alignment of dimeric 4-OT (limon, PDB entry 1OTF) and monomeric *cis*-CaaD (dark salmon, PDB entry 3MF8) to monomeric F4OT (deep teal). **B)** Active site alignment of *cis*-CaaD and F4OT (normal interface). **C)** Active site alignment of 4-OT and F4OT (normal interface).

Fused 4-OT exhibits significant phenylenolpyruvate and 2-HM activity (1.9×10^5 $M^{-1}s^{-1}$ and 2.8×10^6 $M^{-1}s^{-1}$, respectively, Baas *et al.*, *In* Preparation, **Table 4-6**). The key active site residues Pro-1, Arg-11′, Phe-50′, and Arg-39″ of canonical 4-OT are conserved in Fused 4-OT (equivalent to Pro-1, Arg-76, Phe-115, and Arg-104′ in Interface B, **Figure 4-17C**). Arg-104′ is replaced by Arg-39′ at both sites in interface A. Pro-66′ of Fused 4-OT is equivalent to Pro-1″ of the canonical 4-OT, suggesting that the evolution of Fused 4-OT involved gene duplication and fusion of an ancient 4-OT-like gene.

In Interface B, the side chain of Arg-39 (Nη) interacts with the backbone of the linker between the β-α-β units and blocks the side chain of Arg-104′ from interacting with the side chain hydroxyl group of Thr-2 (**Figure 4-7D**). Although 4-OT does not have a linker, the positioning of these arginines resembles that of Arg-39″ and Arg-39 in 4-OT (**Figure 4-17C**). The conformation differs with the Arg-39′ residues in Interface A of Fused 4-OT (**Figure 4-7C**), where the side chain of Arg-39′ (Nη) hydrogen bonds to the side chain hydroxyl group of Thr-2. In 4-OT, Thr-2 is replaced with Ile-2, so a similar interaction cannot be made with Arg-39″.

Table 4-6 Steady-state parameters for Ps01740, Pt0534, Fused 4-OT (Baas BJ, *In Preparation*), *cis*-CaaD, and 4-OT.

						
Enzyme	**k_{cat} (s^{-1})**	**K_M (μM)**	**k_{cat}/K_M (M^{-1} s^{-1})**	**k_{cat} (s^{-1})**	**K_M (μM)**	**k_{cat}/K_M (M^{-1} s^{-1})**
cis-CaaD	N.D.	N.D.	N.D.	0.20 ± 0.03	110 ± 30	1.8×10^3
Linker 1	N.D.	N.D.	N.D.	80 ± 9	230 ± 40	3.5×10^5
Linker 2	2800 ± 130	80 ± 9	3.5×10^7	920 ± 48	685 ± 68	1.3×10^6
Fused 4-OT	580 ± 25	208 ± 16	2.8×10^6	136 ± 8	720 ± 75	1.9×10^5
4-OT	4000 ± 182	62 ± 8	6.5×10^7	73 ± 6	199 ± 23	3.7×10^5

N.D.: Not Detectable

The structure of Fused 4-OT in complex with phenylpyruvate (F4OT-PP) reveals that the ligand is bound at Interface A in two orientations (**Figure 4-18A**), one of which is similar to that observed in a modeling study with the heterohexameric 4-OT shown in **Figure 4-18B** (Burks *et al.*, 2010). However, the hydrogen bonding interactions with the pyruvate moiety differ. In F4OT-PP, the carboxylate is hydrogen bonding to the prolyl nitrogen of Pro-1 and the backbone carbonyl oxygen of Ile-71 in (**Figure 4-18A**), whereas the carboxylate interacts with the side chain of αArg-12 (Nη) in heterohexameric 4-OT (**Figure 4-18B**). In F4OT-PP, the C-2 oxygen of phenylpyruvate interacts with the side chain of Arg-76 (Nη) (**Figure 4-18A**), which is equivalent to αArg-12 in heterohexameric 4-OT.

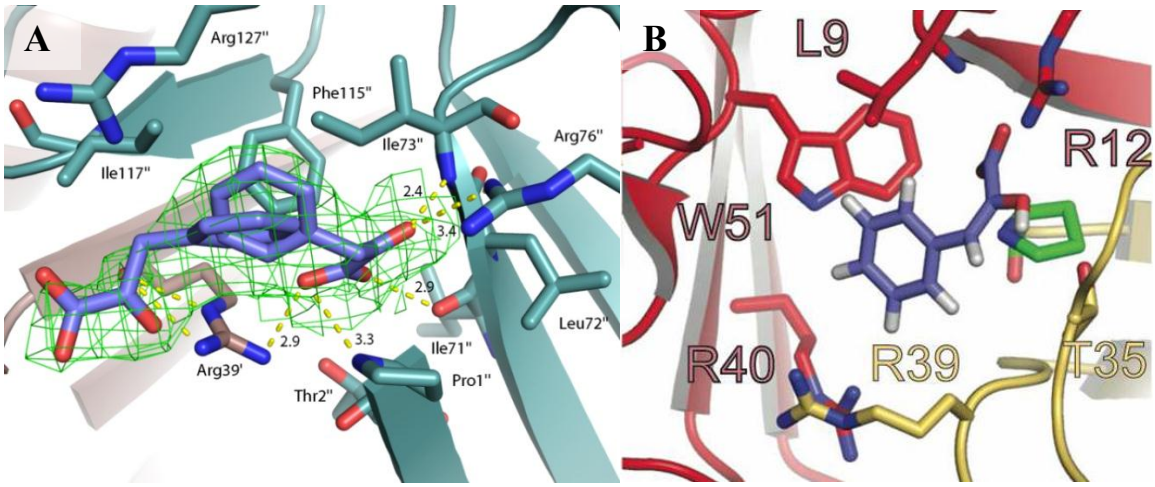


Figure 4-18 Comparison of phenylpyruvate bound to **A)** Fused 4-OT (F4OT-PP) and **B)** heterohexamer 4-OT (Burks *et al.*, 2010).

4.4.4 2-HM Tautomerization in Interface A and B of Fused 4-OT

In F4OT-2HM, the C-1 and C-6 carboxylate groups of 2-HM are bound to side chain of Arg-76 (Nη) and Arg-104′/39′ (Nη), which is flipped 180° relative to the reported orientation of 2-HM in canonical 4-OT, where the C-1 and C-6 carboxylate groups are bound to the equivalent side chains of Arg-11′ (Nη) and Arg-39″ (Nη), respectively (Harris *et al.*, 1999). To investigate the cause of the reverse orientation, the α-carbon atoms of Pro-1, Arg-104′, and Arg-39 in Interface B were aligned to equivalent residues of 4-OT (**Figure 4-19**). In this superposition, Phe-115 (Cα) in Fused 4-OT is shifted 1.6 Å further from Pro-1 than Phe-50′ (Cα) in 4-OT, and Arg-76 (Cα) in Fused 4-OT is shifted 2.8 Å closer to Pro-1 than Arg-11′ in 4-OT. This shift is present in each interface, as the backbone in each active site is superimposable (**Figure 4-7**).

Additionally, the C-2 oxygen of 2-HM in Interface A and B of F4OT-2HM is bonded to a water molecule stabilized by the side chain hydroxyl group of Thr-2 (**Figure 4-8C,E**). In 4-OT, Thr-2 is substituted for Ile-2, and no hydrogen bond is made with the C-2 oxygen. The lack of Thr-2 and the increased distance of Arg-11′ from Pro-1 in 4-OT might cause the 2-HM orientation to be opposite to that observed in F4OT-2HM.

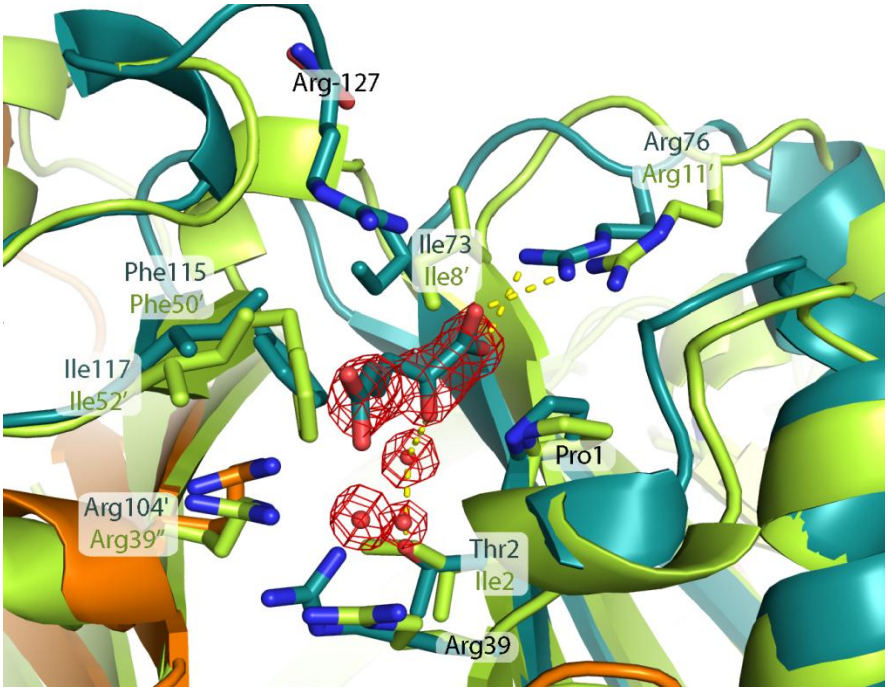


Figure 4-19 Active site superposition of F4OT-2HM (orange and deep teal) and 4-OT (limon, PDB entry 1OTF).

The interaction of the Thr-2 side chain hydroxyl group and Arg-76 (Nη) with 2-HM in Interface A matches interactions in Interface B, but interactions between arginine residues and the C-6 carboxylate groups differ (**Figure 4-10A**). The absence of 2-HM in

one active site of Interface A is explained by the alternate orientation of Arg-76 in that active site which extends away from Pro-1.

Fused 4-OT mutants were prepared and characterized as described elsewhere (Baas BJ, *et al. In Preparation*). The steady state kinetic parameters are shown in **Table 4-7**. The k_{cat}/K_M value for the P1A mutant decreases 200-fold, a drop less extreme than other mutants, of which the R76A mutant was the most impaired (k_{cat}/K_M decreases 770-fold). The relative drop in activity was the same as the decrease for the equivalent 4-OT mutant (R11A, 770-fold) (Harris, *et al.,* 1999). The k_{cat}/K_M value for R39A mutant of Fused 4-OT mutant dropped 330-fold relative to wild-type, which is similar to the decrease for the comparable 4-OT mutant (200-fold). The k_{cat} value for the R104A mutant of Fused 4-OT dropped ~2-fold, and the K_M value did not change. Arg-104 is only present in Interface B, so the data suggest it is required for catalysis in this interface. Arg-104 must be accompanied by Arg-39, though, as the R39A mutant shows its role cannot be substituted in 2-HM tautomerization (330-fold decrease in k_{cat}/K_M). The notion is in agreement with the structure of 2-HM bound in the Interface B — both Arg-39 and Arg-104′ contact the C-6 carboxylate (**Figure 4-8D,E**).

Table 4-7 Steady-state Kinetic Parameters of Fused 4-OT Mutants (Baas BJ, *In Preparation*).

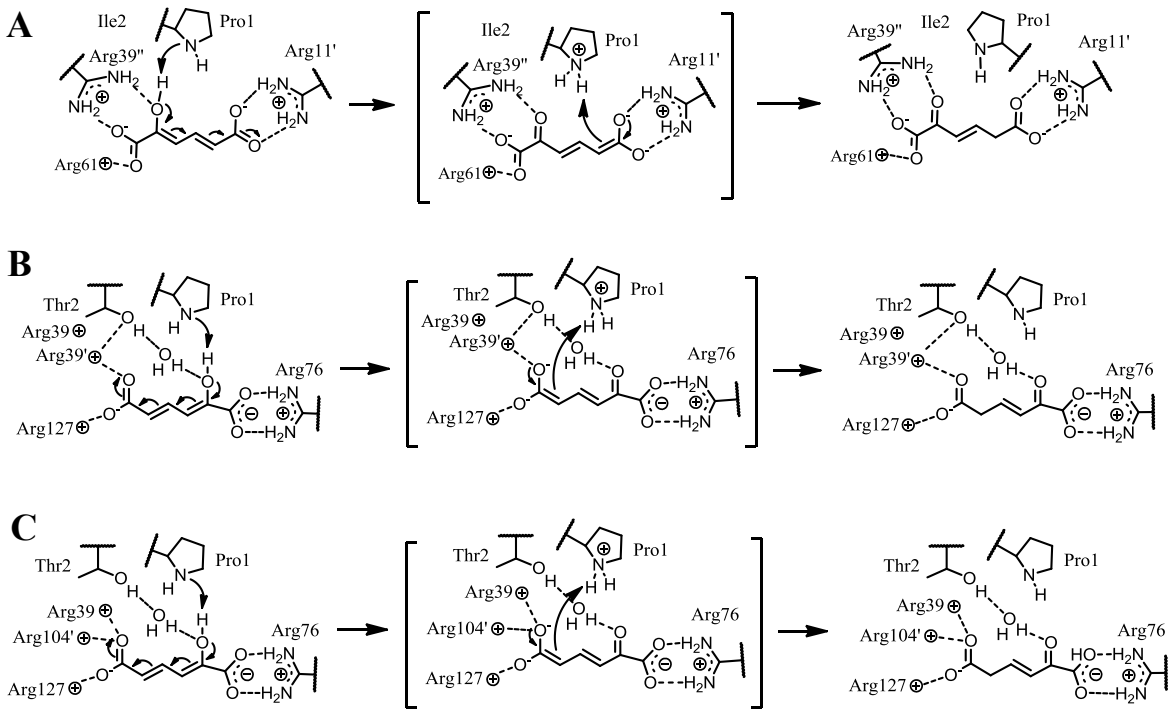
Enzyme	k_{cat} (s^{-1})	rel. k_{cat}	K_m (μM)	rel. K_m	k_{cat}/K_m (M^{-1} s^{-1})	rel. k_{cat}/K_m
WT	580 ± 25	1.0	208 ± 16	1.0	2.79×10^6	1.0
P1A	3.4 ± 0.1	5.9×10^{-3}	221 ± 11	1.06	1.54×10^4	5.5×10^{-3}
R39A	N.D.	N.D.	N.D.	N.D.	8.34×10^3	3.0×10^{-3}
R76A	N.D.	N.D.	N.D.	N.D.	3.74×10^3	1.3×10^{-3}
R104A	313 ± 15	0.54	213 ± 18	1.02	1.47×10^6	0.53
R127A	N.D.	N.D.	N.D.	N.D.	2.84×10^4	1.0×10^{-2}

N.D.: Not Detectable

Surprisingly, the k_{cat}/K_M value for the R127A mutant was 100-fold less than that of wild-type. This result is contrary to the the result seen with 4-OT where the K_M for the R61A mutant increased ~2-fold (Harris *et al.*, 1998), but is more consistent with the study of TomN (Burks, *et al.*, 2011). The decrease in activity for the R127A mutant of Fused 4-OT and distances of the side chain of Arg-127 (Nη) to the C-6 carboxylate (3.7 Å in the Interface A and 4.2 Å in the Interface B) suggest that the C-6 carboxylate may be stabilized by Arg-127 in Fused 4-OT during enediolate formation and collapse. Conversely, intermediates in 2-HM tautomerization in 4-OT are stabilized by Arg-11′, as this residue binds the C-6 carboxylate, and Arg-61 (equivalent to Arg-127) is not nearby to participate in catalysis by an interaction with the C-6 carboxylate.

Mechanisms for 2-HM tautomerization in Interfaces A and B are proposed (**Scheme 4-2B,C**, respectively). In both active sites, the prolyl nitrogen of Pro-1 removes the proton of the C-2 hydroxyl group, with the assistance of a water molecule that is

stabilized by the side chain group of Thr-2. Electron rearrangement occurs to form an enediolate species. In the Interface A (**Scheme 4-2B**), the side chain of Arg-39′ (respectively $N\eta_1$ and $N\eta_2$) is bound to the side chain hydroxyl group of Thr-2 and stabilizes one oxygen of the enediolate intermediate. The side chain of Arg-127 stabilizes the second oxygen of the enediolate, and protonation by the nitrogen of Pro-1 occurs at C-5. In the Interface B (**Scheme 4-2C**), one oxygen of the enediolate is stabilized by the side chains of Arg-39 (Nη) and Arg-104′ (Nη), while the other is stabilized by the side chain of Arg-127 (Nη).



Scheme 4-2 Proposed mechanisms for **A)** 4-OT **B)** Fused 4-OT (Interface A) **C)** Fused 4-OT (Interface B).

4.4.5 Structure and Function Relationships of Linkers

Our studies of Ps01740, Pt0534, and Fused 4-OT, contemporary enzymes that link 4-OT and *cis*-CaaD subgroups via sequence identity, have affirmed the consequence of arginine side chain positioning in the processing of monoacid and diacid substrates. Members of the 4-OT subgroup that bind dicarboxylate substrates such as 2-HM (from the meta-fission pathway) have a pair of arginine residues that are on opposite sides of Pro-1 (Arg-11′ and Arg-39″, in 4-OT). Such is the case with Pt0534 and Fused 4-OT, which have efficient tautomerase activity using 2-HM. With Fused 4-OT, one monomer is inverted relative to the others, but each active site retains Arg76 and Arg-39′/Arg-104′ (Arg-11′ and Arg-39″ in 4-OT) for binding 2-HM. This inversion of one subunit and the opposite 2-HM orientation to that proposed for 4-OT (Harris TK, *et al.* 1999) may be anomalies in the 4-OT subgroup, but further investigation is required to determine whether these traits are widespread. Fused 4-OT has a proline at position 65, which may be a relic of a catalytic Pro-1 of an ancestral 4-OT gene. Pt0534 is also a member of the 4-OT subgroup and has Arg-71 and Arg-99′ conserved for binding 2-HM (equivalent to Arg-11′ and Arg-39″ in 4-OT). The 4-OT activity of Pt0534 suggests that Phe-50′ (4-OT numbering) can be replaced by other large hydrophobic residues (Met-110 in Pt0534), as seen before with heterohexamer 4-OT and TomN (Burks *et al.*, 2010, Burks, *et al.*, 2011). The structures of Pt0534-2OP and F4OT-2HM, in addition to low activity of the R127A mutant of Fused 4-OT (and R61A TomN, Burks EA, *et al.*, 2011) indicate that Arg-127 might be important in the 4-OT subgroup.

Members of the *cis*-CaaD subgroup retain Pro-1 and Arg-11′, but lose Arg-39″. The absence of this residue in Ps01740 is consistent its inability to process the dicarboxylate substrate 2-HM. Many other residues that are key for *cis*-CaaD activity have been acquired in Ps01740 (Arg-70, Glu-114, and His-28, *cis*-CaaD numbering) except for Tyr-103, which might explain the absence of *cis*-CaaD activity. The absence of Tyr-103′ allows for the "in" conformation of the 2-oxo-3-propenoate adduct in Ps01740-2OP (**Figure 4-3**), and may contribute to the low hydatase activity relative to other *cis*-CaaD homologues. Covalent modification by 2-OP at Pro-1 in Ps01740 and BP4401 (**Figure 5-2**), both of which catalyze 2-OP hydration, support a covalent mechanism for these enzymes, which might reflect their unknown biological activity. Phenylpyruvate tautomerase activity is exhibited in each linker and functionally connects the 4-OT and *cis*-CaaD subgroups. The anomolies discovered for the linker enzymes show the diversity of structure and behavior within an enzyme subgroup and could not be predicted by sequence alone.

Chapter 5: The Malonate Semialdehyde Decarboxylase Homologues BP4401 and YusQ

5.1 INTRODUCTION

The enzymes designated BP4401 from *Burkholderia phymatum* strain STM815 and YusQ from *Bacillus substilis*, respectively share 26 and 29% sequence identity with malonate semialdehyde decarboxylase (MSAD) from *Pseudomonas pavonaceae* 170. The latter enzyme participates in the 1,3-dichloropropene degradation pathway (**Section 3.5**), whereas the homologues have unknown functions. These homologues are of interest because they possess the core catalytic residues of MSAD (Pro-1, Asp-37, and Arg-76), but do not have decarboxylase activity using malonate semialdehyde. Bp4401 shows increased hydratase activity using 2-oxo-3-pentynoate (2-OP, 3000 $M^{-1}s^{-1}$) relative to MSAD (600-1100 $M^{-1}s^{-1}$), while YusQ has moderate hydratase activity (440 $M^{-1}s^{-1}$) (Huddleston *et al.*, 2014). BP4401 and YusQ are not encoded near genes for 1,3-dichloropropene degradation, and biological roles for the enzymes might be related to their hydratase activity. In this chapter, crystallography is used to provide clues to establish the determinants of hydratase activity, to identify "missing" determinants of decaroxylase activity, and to assist identification of biological substrates.

5.2 EXPERIMENTAL PROCEDURES

Crystallization and Crystal Soaking of Bp4401. The diffraction-quality crystals of Bp4401 were obtained in PEG 3350 (22-26%) with ZnAc (0.2 M). To obtain the structure of BP4401 inactivated by 2-oxo-3-pentynoate (2-OP, **BP4401-2OP**) an apo crystal of Bp4401 was soaked in 5.9 mM 2-OP for ~5 min. To obtain the structure of BP4401 inactivated by 3-bromopropiolate (**BP4401-BrP**), an apo crystal of Bp4401 was soaked in 10 µM 3-bromopropiolate for 75 min.

Crystallization and Crystal Soaking of YusQ. The diffraction-quality crystals of YusQ were obtained in 0.1 M MES (pH 6.5-7.0) and 26% PEG 3350. To obtain the the structure of YusQ inactivated by 3-bromopropiolate (**YusQ-BrP**), an apo crystal of YusQ was soaked in reservoir solution containing 3-bromopropiolate (10 µM) for 95 min. Crystals were cryoprotected by 25% glycerol in mother liquor prior to vitrification in liquid nitrogen.

Data Collection and Struture Determination. X-ray diffraction data for BP4401, BP4401-2OP, and YusQ were collected at Advanced Light Source beamline 5.0.3 (ALS, Berkeley, CA). Datasets used for the BP4401-BrP and YusQ-BrP was conducted at ALS beamline 5.0.1. The datasets were indexed, integrated, and scaled using HKL2000 (Otwinowski & Minor, 1997). Statistics for data collection are summarized in **Table 2-9** and **Table 5-1**. All structures were determined by molecular replacement using Phaser (McCoy, 2007) and Autobuild (Terwilliger *et al.*, 2008) from PHENIX (Adams et al., 2004). The native structure of YusQ (**YusQ**) was determines by molecular replacement

using a crystal structure of MSAD FG41 (a homologue isolated from *Cornyeform* strain FG41) inactivated by 3-chloropropiolate (29.6% sequence identity, PDB entry 3N4G) as a search model. The YusQ structure was used as a search model for YusQ-BrP, BP4401, and BP4401-BrP. A monomer of BP4401 was used as a search model for the solution of BP4401-2OP. Molecular replacement was performed using Phaser (McCoy, 2007) and Autobuild (Terwilliger *et al.*, 2008). Structures were refined using Phenix Refine (Afonine et al., 2012). Manual model building was done using COOT (Emsley *et al.*, 2010). Model structures were evaluated during and after refinement using Molprobity (Chen *et al.*, 2010). The statisitics for data collection and refinement is summarized in **Table 5-1**.

5.3 RESULTS

BP4401 crystallized in the $P12_11$ and $P2_12_12_1$ space groups with six and three monomers per asymmetric unit, respectively. The native BP4401 crystal structure (**BP4401**) was solved to 2.2 Å via molecular replacement using YusQ as a search model. 125 (of 126 amino acids in Bp4401) were resolved in 5 monomers, and all were resolved in the sixth monomer.

YusQ crystallized in the $P2_13$ space group with one monomer per asymmetric unit. The 2.1 Å resolution crystal structure of native YusQ (**YusQ**) was determined by molecular replacement using inactivated FG41 MSAD as a search model. The N-terminal 125 residues of the 127 amino acid protein were resolved.

BP4401 and YusQ fold into two β-α-β units, as expected for long TSF members. The monomers trimerize and possess three-fold rotational symmetry. Short beta strands in the linker and C-terminus interact with the neighboring monomer.

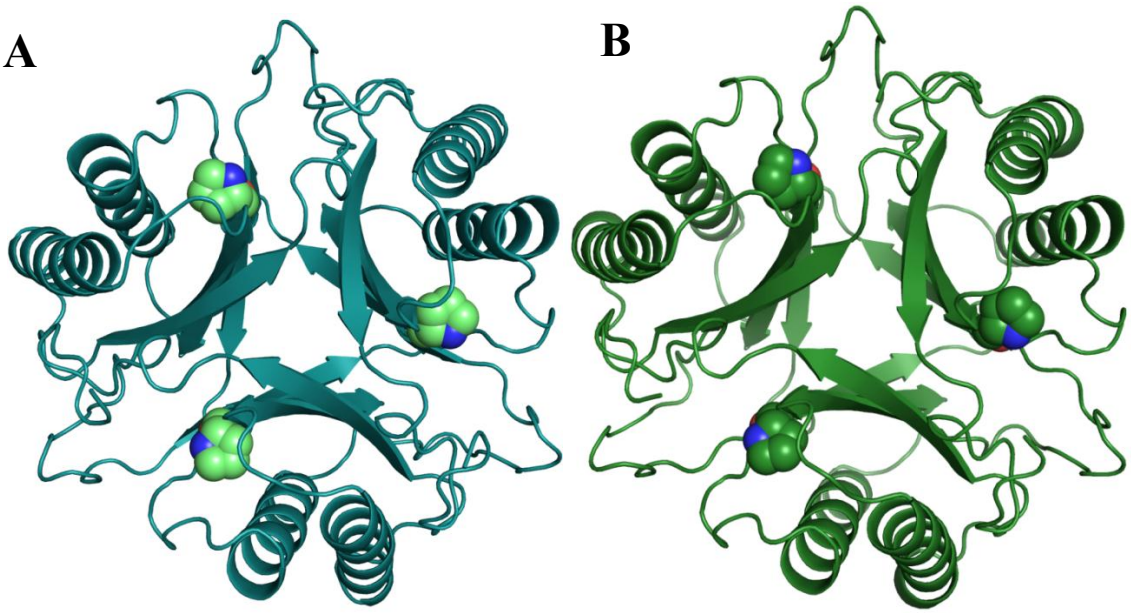


Figure 5-1 The 2.2 Å resolution BP4401 crystal structure and the 2.1 Å resolution YusQ crystal structure.

Pro-1 atoms are shown as spheres. **A)** BP4401 hexamer. **B)** YusQ hexamer generated by a symmetry operation.

The crystal structure of BP4401 inactivated by 2-OP (**BP4401-2OP**) was solved to 1.9 Å resolution using a soaking procedure. All 126 residues were resolved via molecular replacement using native BP4401 as a search model. Three monomers were present in each asymmetric unit. The $2F_o$-F_c electron density map revealed 2-OP had modified the prolyl nitrogen at Pro-1 in two active sites and was converted to

acetopyruvate in the third with no covalent modification of the prolyl nitrogen (**Figure 5-2**). In each case, the C-1 carboxylate of the 2-oxo-3-pentenoate adduct was bound to the side chain of Arg-76 (Nη). The position of the C-2 oxygen atom varies in each active site near Thr-73. The side chain hydroxyl group of Tyr-124 hydrogen bonds to the side chain carboxylate group of Asp-37 in one active site (**Figure 5-2B**), but not in the others (**Figure 5-2A,C**). The C-4 oxygen of acetopyruvate forms a hydrogen bond with the side chain carboxylate oxygen of Asp-37 (**Figure 5-2A**).

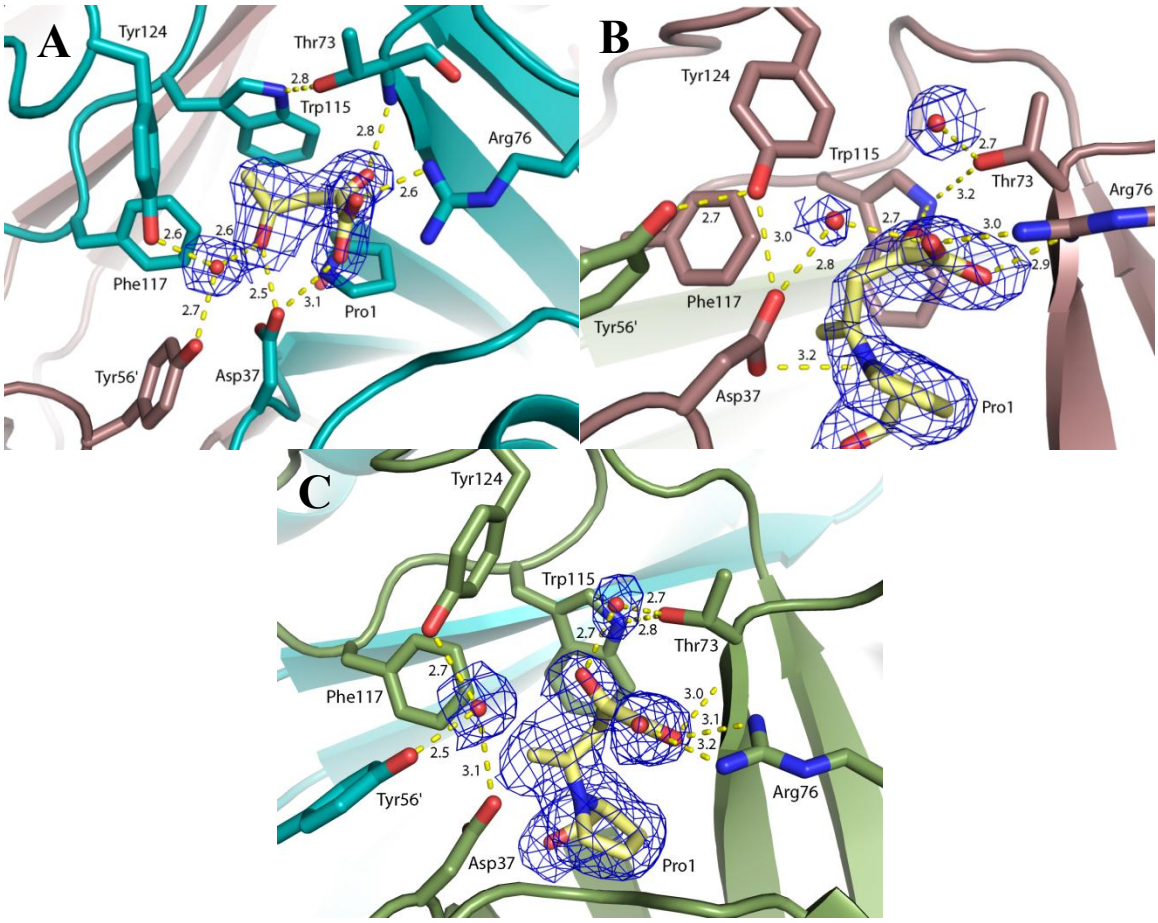
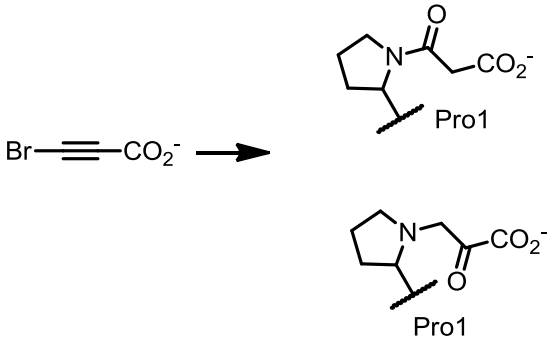


Figure 5-2 Active sites of the 1.9 Å resolution structure of BP4401 inactivated by 2-OP. $2F_o$-F_c electron density of ligands and nearby waters are contoured at 0.5 σ. Active sites contain **A)** acetopyruvate and **B, C)** 2-oxo-3-pentenoate adducts.

Bp4401 and YusQ inactivated by 3-bromopropiolate (**BP4401-BrP** and **YusQ-BrP**) using a soaking procedure and were solved to 1.8 and 2.8 Å resolution, respectively. Molecular replacement was used to solve BP4401-BrP and YusQ-BrP using the native YusQ as a search model. BP4401-BrP and YusQ-BrP contained six and one monomers per asymmetric unit, respectively.

The initial $2F_o$-F_c electron density map in the YusQ-BrP structure in complex inactivated by 3-bromopropiolate suggested a 2-oxopropanoate adduct at Pro-1 (**Scheme 5-1**). The true ligand may be 3-oxopropanoate, since there is precedence for this adduct (Poelarends *et al.*, 2005, Almrud *et al.*, 2005) and the resolution of the YusQ-BrP structure is low (2.8 Å).



Scheme 5-1 Possible BP4401 and YusQ modifications at Pro-1 (prolyl nitrogen) using 3-bromopropiolate.

Hydration at C-3 produces the 3-Oxopropanoate (**top**) adduct and water addition at C-2 allows for the 2-oxopropaonate adduct (**bottom**).

The $2F_o$-F_c electron density map of BP4401-BrP structure show density near Pro-1 in all six monomers larger than the 3-oxo-propanoate adduct expected. Zinc was modeled in the active site (**Figure 5-3A**), given that a high concentration of zinc was present in the reservoir buffer (0.2 M). In addition, the bond lengths (1.9, 2.0, 2.3, and 2.6 Å) are comparable with similar zinc compounds with Zn-O bond lengths of 2.1 Å (Chen & Powers, 1995). The C-1 carboxylate oxygen atoms of the 2-oxo-propanoate adduct in

BP4401-BrP and YusQ-BrP are bound to the side chain of Arg-76 (Nη). In YusQ, the 2-oxopropanoate ligand was modelled in the structure instead of the 3-propanoate adduct due to the better fit in the $2F_o$-F_c map.

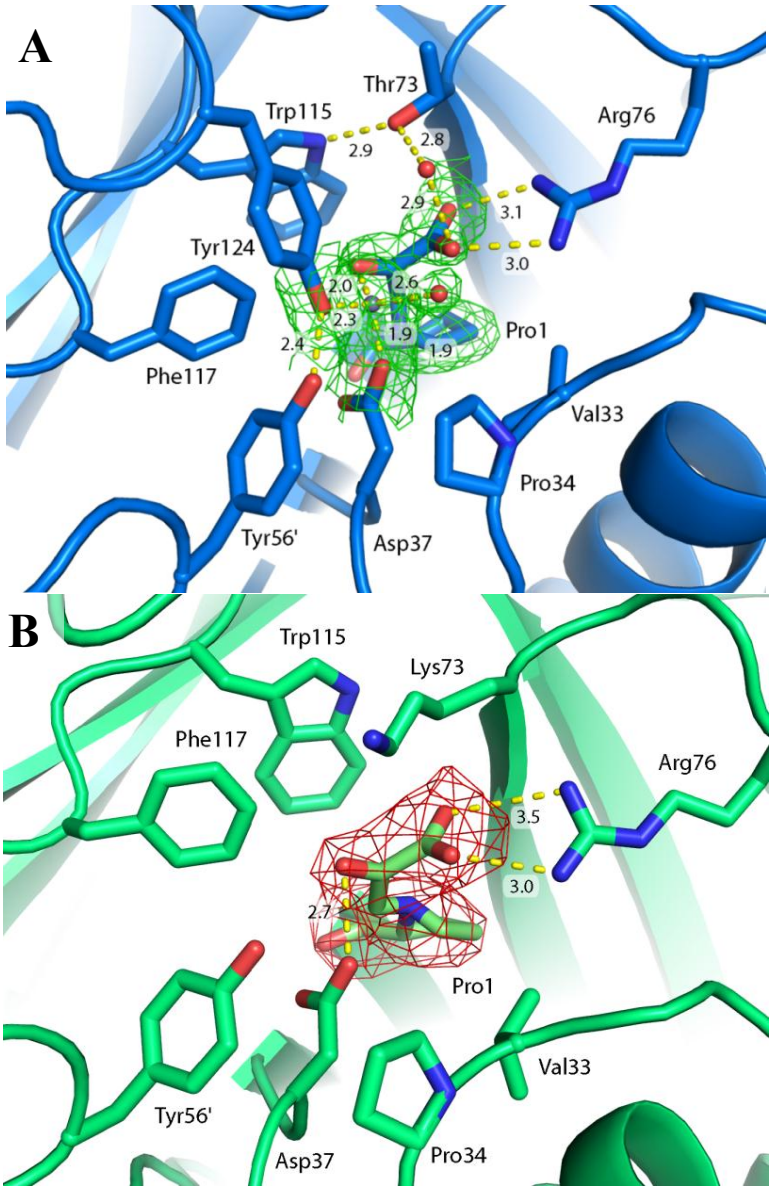


Figure 5-3 Active sites of the 1.8 Å resolution BP4401-BrP structure and the 2.8 Å YusQ-BrP structure.

A) $2F_o$-F_c electron density map surrounding Pro-1 and the 2-oxopropanoate adduct, zinc (grey sphere), and water (red sphere) is contoured at 0.5 σ. **B)** $2F_o$-F_c continuous density surrounding Pro-1 and 2-oxopropanoate adduct is contoured at 1.0 σ.

Data Collection	BP4401	BP4401-2-OP	BP4401-BrP	YusQ	YusQ-BrP
space group	$P12_11$	$P2_12_12_1$	$P12_11$	$P2_13$	$P2_13$
cell dimensions					
a, b, c (Å)	53.2, 136.6, 55.4	65.0, 67.9, 100.5	53.3, 136.4, 55.6	84.8, 84.8, 84.8	83.2, 83.2, 83.2
α, β, γ (deg)	90.0, 105.9, 90.0	90.0, 90.0, 90.0	90.0, 106.4, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
resolution (Å)	50.00 - 2.18	50.00 - 1.88	50.00 - 1.80	50.00 - 2.14	50.00 - 2.81
R_{sym} (%) (linear R factor)	0.097 (0.388)	0.096 (0.622)	0.256 (0.068)	0.400 (0.975)	0.127 (0.971)
$I/\sigma I$	15 (2.3)	21.8 (1.8)	20.4 (2.8)	21 (3.1)	16.9 (1.2)
completeness (%)	99.9 (100.0)	99.1 (99.7)	100.0 (100.0)	100.0 (100.0)	99.9 (100.0)
Refinement					
resolution (Å)	49.65 - 2.19	40.40 - 1.88	49.68 - 1.80	42.41 - 2.11	48.042 -2.80
no. of reflections	39590	36524	70284	11477	4868
R_{work}/R_{free} (%)	18.68/23.23	19.01/22.11	14.75/18.79	17.17/19.97	24.65/29.18
no. of atoms					
protein	6120	3111	6183	1052	1057
water	441	324	892	98	2
Average *B* factor (Å^2)					
protein	35.6	38.6	21.9	29.5	63.1
water	39.3	43.6	35.5	38.0	23.1
rmsd					
bond lengths (Å)	0.015	0.005	1.748	0.008	0.007
bond angles (deg)	1.252	0.980	1.665	0.968	1.240
Ramachandran plot (%)					
favored regions	97.16	97.04	97.00	95.90	95.87
allowed regions	2.84	2.96	2.86	4.10	4.13
disallowed regions	0.00	0.00	0.14	0.00	3.30

Data for the last resolution shell are given in parentheses.

Table 5-1 Crystallographic Data and Refinement statistics for BP4401, BP4401-2OP, BP4401-BrP, YusQ, and YusQ-BrP.

5.4 DISCUSSION

The tertiary structures of YusQ, BP4401, and MSAD are comparable (**Figure 5-4**), as RMSD values of the proteins range between 0.55 and 0.72. MSAD deviates from YusQ and BP4401 at the loop in the β-hairpin in the linker region. Additionally, MSAD possesses an α-helix at its C-terminus, which is not present in YusQ or BP4401.

Figure 5-4 Superposition of MSAD (deep salmon, PDB entry 2AAG), BP4401 (deep teal), and YusQ (forest) monomers.

Asp-37 has been proposed to activate water for addition to 2-OP in FG41 MSAD, (38% sequence ID with MSAD, referred to as **FG41**). Given the aspartate residue is conserved in MSAD, BP4401, and YusQ (in addition to Pro-1 and Arg-76, MSAD

numbering), it is not surprising that each of these enzymes has hydration activity with 2-OP. The k_{cat}/K_M values for FG41, MSAD, YusQ, and BP4401 are 20, 600-1100, 440, and 3000 $M^{-1}s^{-1}$, respectively (Huddleston *et al.*, 2014). The crystal structure of BP4401-2OP is consistent with the role of Asp-37 in water activation as the C-4 oxygen of acetopyruvate ligand hydrogen bonds with the side chain carboxylate of Asp-37 (**Figure 5-2A**). The 2-oxo-3-pentenoate adduct (**Figure 5-2B,C**) supports a covalent mechanism for hydration as noted for Ps01740, the *cis*-CaaD homologue with low level hydration activity, that was shown to be modified by 2-OP in the crystal structure (**Section 4.4.1**)

The carboxylate groups of the propanoate adducts (**Figure 5-5**) of the 3-halopropiolate-inactivated structures of MSAD and YusQ are oriented towards the side chain of Arg-75 (Nη, MSAD numbering). Though the adducts are oriented similarly, differences in active site composition may prevent YusQ from exhibiting MSAD activity. An obvious difference in YusQ is Lys-73, which replaces Arg-73 in MSAD.

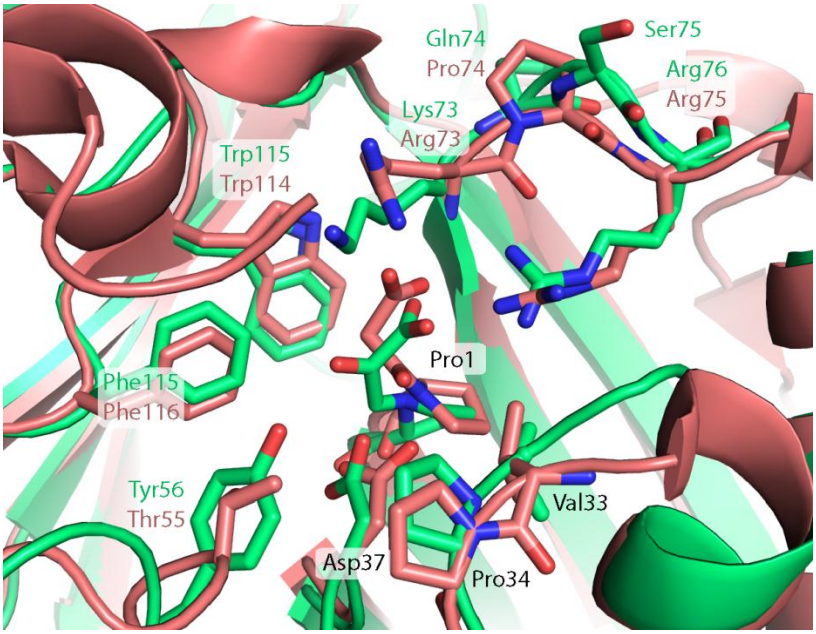


Figure 5-5 Superposition of MSAD with 3-oxopropanoate adduct (deep salmon, PDB entry 2AAL) and YusQ-BrP with 2-oxopropanoate adduct (lime green).

Though key residues Pro-1, Asp-37, and Arg-75 (**Figure 5-6**) in MSAD are conserved in BP4401 and YusQ, neither enzyme exhibits MSAD activity. BP4401 and YusQ lack an equivalent to Arg-73, which is replaced by Thr-73 and Lys-73, respectively. The lack of arginine at this position did not eliminate MSAD activity for FG41, though, when Arg-73 is replaced with Gln-73. The FG41-catalyzed decarboxylation is robust, as MSAD and FG41 have comparable k_{cat}/K_M values at 8700 $M^{-1}s^{-1}$ and 7400 $M^{-1}s^{-1}$ (Huddleston *et al.*, 2014). Mutagenesis and structural studies of FG41 suggest that Thr-72, Gln-73, and Tyr-123 likely replace Arg-73 and Arg-76 for MSAD activity (Guo *et al.*, 2013). Thr-72 and Gln-73 are not conserved in BP4401 or

YusQ, but Tyr-123 is conserved in BP4401. Thr-72 and Gln-73 (FG41 numbering), or Arg-73 (MSAD numbering) may be required to introduce decarboxylase activity in BP4401. These residues in addition to Tyr-123 may be required for MSAD activity in YusQ.

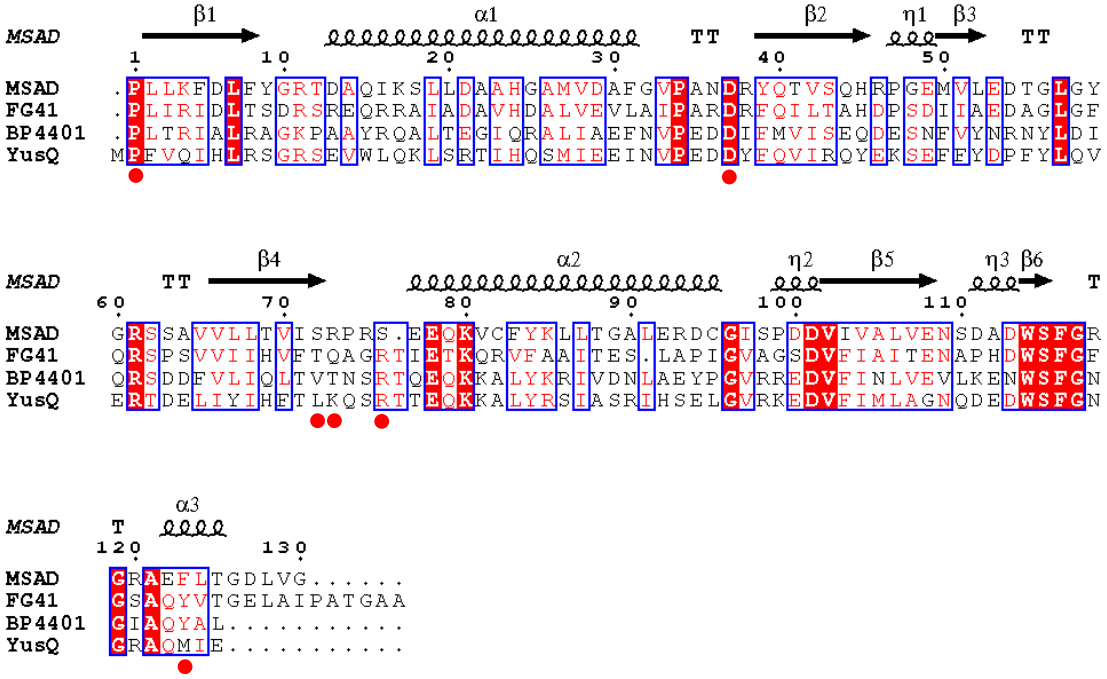


Figure 5-6 Sequence alignment of MSAD, FG41, BP4401, and YusQ.

The residues discussed in the text are indicated with a red dot. Similar residues are in red type. Identical residues are indicated by filled red background. Secondary structure is based on MSAD (PDB entry 2AAG).

References

Adachi K, Iwabuchi T, Sano H & Harayama S (1999) Structure of the ring cleavage product of 1-hydroxy-2-naphthoate, an intermediate of the phenanthrene-degradative pathway of *Nocardioides* sp. strain KP7. *Journal of Bacteriology* **181**: 757-763.

Adams PD, Gopal K, Grosse-Kunstleve RW, *et al.* (2004) Recent developments in the PHENIX software for automated crystallographic structure determination. *Journal of Synchrotron Radiation* **11**: 53-55.

Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH & Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D* **68**: 352-367.

Aisaka K, Igarashi A, Yamaguchi K & Uwajima T (1991) Purification, crystallization and characterization of *N*-acetylneuraminate lyase from *Escherichia coli*. *Biochemical Journal* **276 (Pt 2)**: 541-546.

Almrud JJ, Poelarends GJ, Johnson WH, Serrano H, Hackert ML & Whitman CP (2005) Crystal Structures of the Wild-Type, P1A Mutant, and Inactivated Malonate Semialdehyde Decarboxylase: A Structural Basis for the Decarboxylase and Hydratase Activities†,‡. *Biochemistry* **44**: 14818-14827.

Atkinson SC, Dogovski C, Downton MT, Pearce FG, Reboul CF, Buckle AM, Gerrard JA, Dobson RCJ, Wagner J & Perugini MA (2012) Crystal, Solution and *In silico* Structural Studies of Dihydrodipicolinate Synthase from the Common Grapevine. *PLoS ONE* **7**: e38318.

Atlas RM & Hazen TC (2011) Oil Biodegradation and Bioremediation: A Tale of the Two Worst Spills in U.S. History. *Environmental Science & Technology* **45**: 6709-6715.

Baas BJ, LeVieux JA, Zhang Y, Whitman CP, *In Preparation* Sructure and Kinetic analysis of Fused 4-OT.

Baas BJ, Poddar H, Geertsema EM, Rozeboom HJ, de Vries MP, Permentier HP, Thunnissen A-MWH & Poelarends GJ (2015) Functional and Structural Characterization of an Unusual Cofactor-Independent Oxygenase. *Biochemistry* **54**: 1219-1232.

Baas BJ, Zandvoort E, Wasiel AA, Quax WJ & Poelarends GJ (2011) Characterization of a newly identified mycobacterial tautomerase with promiscuous dehalogenase and hydratase activities reveals a functional link to a recently diverged *cis*-3-chloroacrylic acid dehalogenase. *Biochemistry* **50**: 2889-2899.

Bahnson BJ, Anderson VE & Petsko GA (2002) Structural mechanism of enoyl-CoA hydratase: Three atoms from a single water are added in either an E1cb stepwise or concerted fashion. *Biochemistry* **41**: 2621-2629.

Barbosa JA, Smith BJ, DeGori R, Ooi HC, Marcuccio SM, Campi EM, Jackson WR, Brossmer R, Sommer M & Lawrence MC (2000) Active site modulation in the *N*-acetylneuraminate lyase sub-family as revealed by the structure of the inhibitor-complexed Haemophilus influenzae enzyme. *Journal of Molecular Biology* **303**: 405-421.

Bennett JP, Bertin L, Moulton B, Fairlamb IJ, Brzozowski AM, Walton NJ & Grogan G (2008) A ternary complex of hydroxycinnamoyl-CoA hydratase-lyase (HCHL) with acetyl-CoA and vanillin gives insights into substrate specificity and mechanism. *Biochemical Journal* **414**: 281-289.

Blickling S, Renner C, Laber B, Pohlenz HD, Holak TA & Huber R (1997) Reaction mechanism of *Escherichia coli* dihydrodipicolinate synthase investigated by X-ray crystallography and NMR spectroscopy. *Biochemistry* **36**: 24-33.

Boldrin B, Tiehm A & Fritzsche C (1993) Degradation of phenanthrene, fluorene, fluoranthene, and pyrene by a *Mycobacterium* sp. *Applied and Environmental microbiology* **59**: 1927-1930.

Bolton JL, Trush MA, Penning TM, Dryhurst G & Monks TJ (2000) Role of quinones in toxicology. *Chemical Research in Toxicology* **13**: 135-160.

Bugg TD & Ramaswamy S (2008) Non-heme iron-dependent dioxygenases: unravelling catalytic mechanisms for complex enzymatic oxidations. *Current Opinion in Chemical Biology* **12**: 134-140.

Burks EA, Yan W, Johnson WH, Li W, Schroeder GK, Min, C, Gerrantana B, Zhang Y, Whitman CP (2011) Kinetic, Crystallographic, and Mechanistic Characterization of TomN: Elucidation of a Function for a 4-Oxalocrotonate Tautomerase Homologue in the Tomamycin Biosynthetic Pathway. *Biochemistry* **50:** 7600-7611.

Burks EA, Fleming CD, Mesecar AD, Whitman CP & Pegan SD (2010) Kinetic and Structural Characterization of a Heterohexamer 4-Oxalocrotonate Tautomerase from *Chloroflexus aurantiacus* J-10-fl: Implications for Functional and Structural Diversity in the Tautomerase Superfamily. *Biochemistry* **49**: 5016-5027.

Chauhan A, Fazlurrahman, Oakeshott JG & Jain RK (2008) Bacterial metabolism of polycyclic aromatic hydrocarbons: strategies for bioremediation. *Indian Journal of Microbiology* **48**: 95-113.

Chen CH, Kenyon GL, Curtin F, Harayama S, Bembenek ME, Hajipour G, Whitman CP (1992) 4-Oxalocrotonate Tautomerase, an Enzyme Composed of 62 Amino Acid Residues per Monomer. *The Journal of Biological Chemistry* **267**: 17716-17721.

Chen D & Powers L (1995) Synthesis and structures of Zn(C6H12OS2)2(ClO4)2 and Zn(C3H6NS2)2(C3H4N2)— Model compounds for the Zn sites in RNA polymerase. *Journal of Inorganic Biochemistry* **58**: 245-253.

Chen VB, Arendall WB, III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D* **66**: 12-21.

Choi KH, Lai V, Foster CE, Morris AJ, Tolan DR & Allen KN (2006) New superfamily members identified for Schiff-base enzymes based on verification of catalytically essential residues. *Biochemistry* **45**: 8546-8555.

Coitinho JB, Costa DMA, Guimaraes SL, de Goes AM & Nagem RAP (2012) Expression, purification and preliminary crystallographic studies of NahF, a salicylaldehyde dehydrogenase from *Pseudomonas putida* G7 involved in naphthalene degradation. *Acta Crystallographica Section F* **68**: 93-97.

Coitinho JB, Pereira MS, Costa DMA, Guimaraes SL, Araujo SS, Hengge AC, Brandao TAS & Nagem RAP (2016) Structural and Kinetic Properties of the Aldehyde Dehydrogenase NahF, a Broad Substrate Specificity Enzyme for Aldehyde Oxidation. *Biochemistry* **55**: 5453-5463.

Czerwinski RM, Harris TK, Massiah MA, Mildvan AS & Whitman CP (2001) The structural basis for the perturbed pK(a) of the catalytic base in 4-oxalocrotonate tautomerase: Kinetic and structural effects of mutations of Phe-50. *Biochemistry* **40**: 1984-1995.

Czerwinski RM, Johnson WH, Whitman CP, Harris TK, Abeygunawardana C & Mildvan AS (1997) Kinetic and structural effects of mutations of the catalytic amino-terminal proline in 4-oxalocrotonate tautomerase. *Biochemistry* **36**: 14551-14560.

Czerwinski RM, Harris TK, Johnson WH, Legler PM, Stivers JT, Mildvan AS & Whitman CP (1999) Effects of mutations of the active site arginine residues in 4-oxalocrotonate tautomerase on the pK(a) values of active site residues and on the pH dependence of catalysis. *Biochemistry* **38**: 12358-12366.

Daisy BH, Strobel GA, Castillo U, Ezra D, Sears J, Weaver DK & Runyon JB (2002) Naphthalene, an insect repellent, is produced by Muscodor vitigenus, a novel endophytic fungus. *Microbiology* **148**: 3737-3741.

Daniels AD, Campeotto I, van der Kamp MW, Bolt AH, Trinh CH, Phillips SE, Pearson AR, Nelson A, Mulholland AJ & Berry A (2014) Reaction mechanism of *N*-acetylneuraminic acid lyase revealed by a combination of crystallography, QM/MM simulation, and mutagenesis. *ACS chemical biology* **9**: 1025-1032.

Dayaram YK, Talaue MT, Connell, ND, and Venketraraman V, Characterization of Glutathione Metabolic Mutant of *Mycobacterium tuberculosis* and Its Resistance to Glutathione and Nitroglutathione. *Journal of Bacteriology* **188**: 1364-1372

Davidson R, Baas BJ, Akiva E., Holliday J, LeVieux, J, Zhang, YJ, Whitman, CP, Babbitt, PC (*In Preparation*) A global view of structure-function relationships in the tautomerase superfamily.

Davies JI & Evans WC (1964) Oxidative metabolism of naphthalene by soil pseudomonads. The ring-fission mechanism. *Biochemical Journal* **91**: 251-261.

de Jong RM, Brugman W, Poelarends GJ, Whitman CP & Dijkstra BW (2004) The X-ray structure of *trans*-3-chloroacrylic acid dehalogenase reveals a novel hydration mechanism in the tautomerase superfamily. *Journal of Biological Chemistry* **279**: 11546-11552.

de Jong RM, Bazzacco P, Poelarends GJ, Johnson WH, Kim YJ, Burks EA, Serrano H, Thunnissen A, Whitman CP & Dijkstra BW (2007) Crystal structures of native and inactivated *cis*-3-chloroacrylic acid dehalogenase - Structural basis for substrate specificity and inactivation by (*R*)-oxirane-2-carboxylate. *Journal of Biological Chemistry* **282**: 2440-2449.

Dobson RC, Perugini MA, Jameson GB & Gerrard JA (2009) Specificity versus catalytic potency: The role of threonine 44 in *Escherichia coli* dihydrodipicolinate synthase mediated catalysis. *Biochimie* **91**: 1036-1044.

Dobson RC, Devenish SR, Turner LA, Clifford VR, Pearce FG, Jameson GB, & Gerrard JA (2005) Role of arginine 138 in the catalysis and regulation of *Escherichia coli* dihydrodipicolinate synthase. *Biochemistry* **44**:13007-13013.

Dobson RCJ, Valegård K & Gerrard JA (2004) The Crystal Structure of Three Site-directed Mutants of Escherichia coli Dihydrodipicolinate Synthase: Further Evidence for a Catalytic Triad. *Journal of Molecular Biology* **338**: 329-339.

Dobson RCJ, Griffin MDW, Devenish SRA, Pearce FG, Hutton CA, Gerrard JA, Jameson GB & Perugini MA (2008) Conserved main-chain peptide distortions: A proposed role for Ile203 in catalysis by dihydrodipicolinate synthase. *Protein Science* **17**: 2080-2090.

Douben PET (2003) *PAHs: An Ecotoxicological Perspective*. John Wiley & Sons, Hoboken, NJ.

Eaton RW (2000) *trans-o*-Hydroxybenzylidenepyruvate hydratase-aldolase as a biocatalyst. *Applied and Environmental Microbiology* **66**: 2668-2672.

Eaton RW & Chapman PJ (1992) Bacterial metabolism of naphthalene: construction and use of recombinant bacteria to study ring cleavage of 1,2-dihydroxynaphthalene and subsequent reactions. *Journal of Bacteriology* **174**: 7542-7554.

Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallographica Section D* **66**: 486-501.

Ernest I, Gosteli J, Greengrass CW, Holick W, Jackman DE, Pfaendler HR & Woodward RB (1978) Penems, a new class of beta-lactam antibiotics – 6-acylaminopenem-3-carboxylic acids. *Journal of the American Chemical Society* **100**: 8214-8222.

European Commission (2002) Polycyclic Aromatic Hydrocarbons–Occurrence in foods, dietary exposure and health effects. *Scientific Committee on Food*, Health and consumer Protection Directorate-General.

Evans WC, Fernley HN & Griffiths E (1965) Oxidative Metabolism of Phenanthrene and Anthracene by Soil Pseudomonads. *Biochemical journal* **95**: 819-831.

Ferrara S, Mapelli E, Sello G & Di Gennaro P (2011) Characterization of the aldol condensation activity of the *trans-o*-hydroxybenzylidenepyruvate hydratase-aldolase (tHBP-HA) cloned from *Pseudomonas fluorescens* N3. *Biochimica et biophysica acta* **5**: 622-629.

Griffin MD, Dobson RF, Pearce FG, Antonio L, *et al.* (2008) Evolution of quaternary structure in a homotetrameric enzyme. *Journal of Molecular Biology* **380**: 691-703.

Guo Y, Serrano H, Poelarends GJ, Johnson WH, Hackert ML & Whitman CP (2013) Kinetic, Mutational, and Structural Analysis of Malonate Semialdehyde Decarboxylase from *Coryneform* Bacterium Strain FG41: Mechanistic Implications for the Decarboxylase and Hydratase Activities. *Biochemistry* **52**: 4830-4841.

Habe H & Omori T (2003) Genetics of polycyclic aromatic hydrocarbon metabolism in diverse aerobic bacteria. *Bioscience, Biotechnology, and Biochemistry* **67**: 225-243.

Harayama S, Rekik M, Wasserfallen A & Bairoch A (1987) Evolutionary relationships between catabolic pathways for aromatics – conservation of gene order and nucleotide-sequences of catechol oxidation genes of PWW0 and NAH7 plasmids. *Molecular Genetics and Genomics* **210**: 241-247.

Harris TK, Legler PM, Czerwinski RM, Abeygunawardana C, Massiah MA, Stivers JT, Whitman CP & Mildvan AS (1999) The roles of arginines in the mechanism of 4-oxalocrotonate tautomerase (4-OT). *Biophysical Journal* **76**: 12343-12357.

Harris TK, Czerwinski RM, Johnson WH, Legler PM, Abeygunawardana C, Massiah MA, Stivers JT, Whitman CP & Mildvan AS (1999) Kinetic, stereochemical, and structural effects of mutations of the active site arginine residues in 4-oxalocrotonate tautomerase. *Biochemistry* **38**: 12343-12357.

Hashidoko Y & Tahara S (1998) Stereochemically specific proton transfer in decarboxylation of 4-hydroxycinnamic acids by 4-hydroxycinnamate decarboxylase from *Klebsiella oxytoca*. *Archives of Biochemistry and Biophysics* **359**: 225-230.

Hecht SS (1999) Tobacco Smoke Carcinogens and Lung Cancer. *Journal of the National Cancer Institute* **91**: 1194-1210.

Heine A, DeSantis G, Luz JG, Mitchell M, Wong C-H & Wilson IA (2001) Observation of Covalent Intermediates in an Enzyme Mechanism at Atomic Resolution. *Science* **294**: 369-374.

Heitkamp MA & Cerniglia CE (1988) Mineralization of polycyclic aromatic hydrocarbons by a bacterium isolated from sediment below an oil field. *Applied and Environmental Microbiology* **54**: 1612-1614.

Heitkamp MA & Cerniglia CE (1989) Polycyclic aromatic hydrocarbon degradation by a *Mycobacterium* sp. in microcosms containing sediment and water from a pristine ecosystem. *Applied and Environmental Microbiology* **55**: 1968-1973.

Heitkamp MA, Franklin W & Cerniglia CE (1988) Microbial metabolism of polycyclic aromatic hydrocarbons: isolation and characterization of a pyrene-degrading bacterium. *Applied Environmental Microbiology* **54**: 2549-2555.

Henderson DP, Cotterill IC, Shelton MC & Toone EJ (1998) 2-keto-3-deoxy-6-phosphogalactonate aldolase as a catalyst for stereocontrolled carbon-carbon bond formation. *Journal of Organic Chemistry* **63**: 906-907.

Huddleston JP, Burks EA & Whitman CP (2014) Identification and characterization of new family members in the tautomerase superfamily: Analysis and implications. *Archives of Biochemistry and Biophysics* **564**: 189-196.

Iwabuchi T & Harayama S (1998) Biochemical and genetic characterization of *trans*-2'-carboxybenzalpyruvate hydratase-aldolase from a phenanthrene-degrading *Nocardioides* strain. *Journal of Bacteriology* **180**: 945-949.

Jeffcoat R, Hassall H & Dagley S (1969) Purification and properties of D-4-deoxy-5-oxoglucarate hydro-lyase (decaroxylating). *Biochemical Journal* **115**: 977-&.

Johnson WH, Czerwinski RM, Fitzgerald MC & Whitman CP (1997) Inactivation of 4-Oxalocrotonate Tautomerase by 2-Oxo-3-pentynoate. *Biochemistry* **36**: 15724-15732.

Johnson WH, Hajipour G, Whitman CP (1995) Stereochemical Studies of 5-(Carboxymethyl)-2-hydroxymuconate Isomerase and 5-(Carboxymethyl)-2-oxo-3-hexene-1,6-dioate Decarboxylase from *Escherichia coli C*: Mechanistic and Evolutionary Implications. *Journal of the American Chemical Society* **117**: 8719-8726

Juhasz AL & Naidu R (2000) Bioremediation of high molecular weight polycyclic aromatic hydrocarbons: a review of the microbial degradation of benzo[*a*]pyrene. *International Biodeterioration & Biodegradation* **45**: 57-88.

Karlsson A, Parales JV, Parales RE, Gibson DT, Eklund H & Ramaswamy S (2003) Crystal structure of naphthalene dioxygenase: side-on binding of dioxygen to iron. *Science* **299**: 1039-1042.

Kelley I & Cerniglia CE (1995) Degradation of a mixture of high-molecular-weight polycyclic aromatic hydrocarbons by a *Mycobacterium* strain PYR-1. *Journal of Soil Contamination* **4**: 77-91.

Kelley I, Freeman JP & Cerniglia CE (1990) Identification of metabolites from degradation of naphthalene by a *Mycobacterium* sp. *Biodegradation* **1**: 283-290.

Kelley I, Freeman JP, Evans FE & Cerniglia CE (1993) Identification of metabolites from the degradation of fluoranthene by *Mycobacterium* sp. strain PYR-1. *Applied and Environmental Microbiology* **59**: 800-806.

Kim SJ, Kweon O, Jones RC, Edmondson RD & Cerniglia CE (2008) Genomic analysis of polycyclic aromatic hydrocarbon degradation in *Mycobacterium vanbaalenii* PYR-1. *Biodegradation* **19**: 859-881.

Kim SJ, Kweon O, Jones RC, Freeman JP, Edmondson RD & Cerniglia CE (2007) Complete and integrated pyrene degradation pathway in *Mycobacterium vanbaalenii* PYR-1 based on systems biology. *Journal of Bacteriology* **189**: 464-472.

Kim SJ, Kweon O, Freeman JP, Jones RC, Adjei MD, Jhoo JW, Edmondson RD & Cerniglia CE (2006) Molecular cloning and expression of genes encoding a novel dioxygenase involved in low- and high-molecular-weight polycyclic aromatic hydrocarbon degradation in *Mycobacterium vanbaalenii* PYR-1. *Applied and Environmental Microbiology* **72**: 1045-1054.

Kim Y-H, Engesser K-H & Cerniglia CE (2003) Two polycyclic aromatic hydrocarbon *o*-quinone reductases from a pyrene-degrading *Mycobacterium*. *Archives of Biochemistry and Biophysics* **416**: 209-217.

Kim YH, Freeman JP, Moody JD, Engesser KH & Cerniglia CE (2005) Effects of pH on the degradation of phenanthrene and pyrene by *Mycobacterium vanbaalenii* PYR-1. *Applied Microbiology and Biotechnology* **67**: 275-285.

Kuhm AE, Knackmuss HJ & Stolz A (1993) Purification and properties of 2'-hydroxybenzalpyruvate aldolase from a bacterium that degrades naphthalenesulfonates. *Journal of Biological Chemistry* **268**: 9484-9489.

Kuśmierz M, Oleszczuk P, Kraska P, Pałys E & Andruszczak S (2016) Persistence of polycyclic aromatic hydrocarbons (PAHs) in biochar-amended soil. *Chemosphere* **146**: 272-279.

Kweon O, Kim SJ, Freeman JP, Song J, Baek S & Cerniglia CE (2010) Substrate specificity and structural characteristics of the novel Rieske nonheme iron aromatic ring-hydroxylating oxygenases NidAB and NidA3B3 from *Mycobacterium vanbaalenii* PYR-1. *mBio* **1**:1-11.

Kweon O, Kim SJ, Jones RC, Freeman JP, Adjei MD, Edmondson RD & Cerniglia CE (2007) A polyomic approach to elucidate the fluoranthene-degradative pathway in *Mycobacterium vanbaalenii* PYR-1. *Journal of Bacteriology* **189**: 4635-4647.

Kweon O, Kim SJ, Kim DW, Kim JM, Kim HL, Ahn Y, Sutherland JB & Cerniglia CE (2014) Pleiotropic and Epistatic Behavior of a Ring-Hydroxylating Oxygenase System in the Polycyclic Aromatic Hydrocarbon Metabolic Network from *Mycobacterium vanbaalenii* PYR-1. *Journal of Bacteriology* **196**: 3503-3515.

Kweon O, Kim S-J, Blom J, Kim S-K, Kim B-S, Baek D-H, Park SI, Sutherland JB & Cerniglia CE (2015) Comparative functional pan-genome analyses to build connections between genomic dynamics and phenotypic evolution in polycyclic aromatic hydrocarbon metabolism in the genus *Mycobacterium*. *BMC Evolutionary Biology* **15**: 1-23.

Kweon O, Kim S-J, Holland RD, *et al.* (2011) Polycyclic Aromatic Hydrocarbon Metabolic Network in *Mycobacterium vanbaalenii* PYR-1. *Journal of Bacteriology* **193**: 4326-4337.

Laber B, Gomis-Rüth FX, Romão MJ & Huber R (1992) *Escherichia coli* dihydrodipicolinate synthase. Identification of the active site and crystallization. *Biochemical Journal* **288**: 691-695.

Lawrence MC, Barbosa JA, Smith BJ, Hall NE, Pilling PA, Ooi HC & Marcuccio SM (1997) Structure and mechanism of a sub-family of enzymes related to *N*-acetylneuraminate lyase. *Journal of Molecular Biology* **266**: 381-399.

LeVieux J, Baas, B. J., Whitman, C. P. (*In Preparation*) Purification, Crystallization, and Structural Characterization of Ps01740: a *cis*-3-chloroacrylic Acid Dehalogenase Homologue in *Pseudomonas* sp. UW4.

LeVieux JA, Johnson WH, Erwin K, Li W, Zhang YJ, Whitman CP (2016) The bacterial catabolism of polycyclic aromatic hydrocarbons: Characterization of three hydratase-aldolase-catalyzed reactions. *Perspectives in Science* **9**: 33-41.

Lorentzen E, Siebers B, Hensel R & Pohl E (2005) Mechanism of the Schiff Base Forming Fructose-1,6-bisphosphate Aldolase: Structural Analysis of Reaction Intermediates. *Biochemistry* **44**: 4222-4229.

MacLeod CT & Daugulis AJ (2003) Biodegradation of polycyclic aromatic hydrocarbons in a two-phase partitioning bioreactor in the presence of a bioavailable solvent. *Applied Microbiology and Biotechnology* **62**: 291-296.

McCoy A (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallographica Section D* **63**: 32-41.

Menn FM, Applegate BM & Sayler GS (1993) NAH plasmid-mediated catabolism of anthracene and phenanthrene to naphthoic acids. *Applied and Environmental Microbiology* **59**: 1938-1942.

Moody JD, Freeman JP, Doerge DR & Cerniglia CE (2001) Degradation of phenanthrene and anthracene by cell suspensions of *Mycobacterium* sp. strain PYR-1. *Applied and Environmental Microbiology* **67**: 1476-1483.

Moody JD, Doerge DR, Freeman JP & Cerniglia CE (2002) Degradation of biphenyl by *Mycobacterium* sp strain PYR-1. *Applied Microbiology and Biotechnology* **58**: 364-369.

Moody JD, Freeman JP, Fu PP & Cerniglia CE (2004) Degradation of benzo[a]pyrene by *Mycobacterium vanbaalenii* PYR-1. *Applied and Environmental Microbiology* **70**: 340-345.

Osterman AL, Brooks HB, Jackson L, Abbott JJ & Phillips MA (1999) Lysine-69 plays a key role in catalysis by ornithine decarboxylase through acceleration of the Schiff base formation, decarboxylation, and product release steps. *Biochemistry* **38**: 11814-11826.

Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology,* **276**:307-326. Academic Press.

Patel TR & Gibson DT (1976) Bacterial *cis*-dihydrodiol dehydrogenases: comparison of physicochemical and immunological properties. *Journal of Bacteriology* **128**: 842-850.

Patel TR & Barnsley EA (1980) Naphthalene metabolism by pseudomonads: purification and properties of 1,2-dihydroxynaphthalene oxygenase. *Journal of Bacteriology* **143**: 668-673.

Pearce FG, Dobson RC, Weber A, Lane LA, McCammon MG, Squire MA, Perugini MA, Jameson GB, Robinson CV & Gerrard JA (2008) Mutating the tight-dimer interface of dihydrodipicolinate synthase disrupts the enzyme quaternary structure: toward a monomeric enzyme. *Biochemistry* **47**: 12108-12117.

Peng RH, Xiong AS, Xue Y, Fu XY, Gao F, Zhao W, Tian YS & Yao QH (2008) Microbial biodegradation of polyaromatic hydrocarbons. *FEMS Microbiology Reviews* **32**: 927-955.

Poelarends GJ & Whitman CP (2004) Evolution of enzymatic activity in the tautomerase superfamily: mechanistic and structural studies of the 1,3-dichloropropene catabolic enzymes. *Bioorganic Chemistry* **32**: 376-392.

Poelarends GJ, Saunier R & Janssen DB (2001) *trans*-3-chloroacrylic acid dehalogenase from *Pseudomonas pavonaceae* 170 shares structural and mechanistic similarities with 4-oxalocrotonate tautomerase. *Journal of Bacteriology* **183**: 4269-4277.

Poelarends GJ, Veetil VP & Whitman CP (2008) The chemical versatility of the β-α-β fold: Catalytic promiscuity and divergent evolution in the tautomerase superfamily. *Cell Mol Life Sci* **65**: 3606-3618.

Poelarends GJ, Serrano H, Johnson WH & Whitman CP (2004) Stereospecific Alkylation of *cis*-3-Chloroacrylic Acid Dehalogenase by (*R*)-Oxirane-2-carboxylate: Analysis and Mechanistic Implications†. *Biochemistry* **43**: 7187-7196.

Poelarends GJ, Serrano H, Johnson WH & Whitman CP (2005) Inactivation of Malonate Semialdehyde Decarboxylase by 3-Halopropiolates: Evidence for Hydratase Activity. *Biochemistry* **44**: 9375-9381.

Poelarends GJ, Johnson WH, Serrano H & Whitman CP (2007) Phenylpyruvate tautomerase activity of *trans*-3-chloroacrylic acid dehalogenase: Evidence for an enol intermediate in the dehalogenase reaction? *Biochemistry* **46**: 9596-9604.

Poelarends GJ, Wilkens M, Larkin MJ, van Elsas JD & Janssen DB (1998) Degradation of 1,3-dichloropropene by *Pseudomonas cichorii* 170. *Applied and environmental microbiology* **64**: 2931-2936.

Poelarends GJ, Serrano H, Johnson WH, Jr., Hoffman DW & Whitman CP (2004) The Hydratase Activity of Malonate Semialdehyde Decarboxylase: Mechanistic and Evolutionary Implications. *Journal of the American Chemical Society* **126**: 15658-15659.

Poelarends GJ, Serrano H, Person MD, Johnson WH & Whitman CP (2008) Characterization of Cg10062 from Corynebacterium glutamicum: Implications for the Evolution of *cis*-3-Chloroacrylic Acid Dehalogenase Activity in the Tautomerase Superfamily. *Biochemistry* **47**: 8139-8147.

Poelarends GJ, Serrano H, Person MD, Johnson WH, Murzin AG & Whitman CP (2003) Cloning, Expression, and Characterization of a *cis*-3-Chloroacrylic Acid Dehalogenase: Insights into the Mechanistic, Structural, and Evolutionary Relationship between Isomer-Specific 3-Chloroacrylic Acid Dehalogenases. *Biochemistry* **43**: 759-772.

Ranc B, Faure P, Croze V & Simonnot MO (2016) Selection of oxidant doses for in situ chemical oxidation of soils contaminated by polycyclic aromatic hydrocarbons (PAHs): A review. *Journal of Hazardous Materials* **312**: 280-297.

Reimer M (1931) Benzalpyruvic acid dibromide. *Journal of the American Chemical Society* **53**: 3147-3149.

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A Laboratory Manual, 2nd ed.,. *Cold Spring Harbor Laboratory Press, Plainview, NY*.

Sareen D, Newton GL, Fahey RC, and Buchmeier NA (2003) Mycothiol is essential for growth of *Mycobacterium tuberculosis* Erdman. *Journal of Bacteriology* **185**: 6736-6740.

Sello G & Di Gennaro P (2013) Aldol reactions of the *trans-o*-hydroxybenzylidenepyruvate hydratase-aldolase (tHBP-HA) from *Pseudomonas fluorescens* N3. *Applied biochemistry and biotechnology* **170**: 1702-1712.

Sisler FD & Zobell CE (1947) Microbial utilization of carcinogenic hydrocarbons. *Science* **106**: 521-522.

Smith BJ, Lawrence MC & Barbosa JARG (1999) Substrate-Assisted Catalysis in Sialic Acid Aldolase. *The Journal of Organic Chemistry* **64**: 945-949.

Sota M, Yano H, Ono A, Miyazaki R, Ishii H, Genka H, Top EM & Tsuda M (2006) Genomic and Functional Analysis of the IncP-9 Naphthalene-Catabolic Plasmid NAH7

and Its Transposon Tn4655 Suggests Catabolic Gene Spread by a Tyrosine Recombinase. *Journal of Bacteriology* **188**: 4057-4067.

Stingley RL, Brezna B, Khan AA, Cerniglia CE (2004) Novel organization of genes in a phthalate degradation operon of *Mycobacterium vanbaalenii* PYR-1. *Microbiology* **150**: 3749-3761.

Stingley RL, Khan AA & Cerniglia CE (2004) Molecular characterization of a phenanthrene degradation pathway in *Mycobacterium vanbaalenii* PYR-1. *Biochemical and Biophysical Research Communications* **322**: 133-146.

Stivers JT, Abeygunawardana C, Mildvan AS, Hajipour G & Whitman CP (1996) 4-Oxalocrotonate Tautomerase: pH Dependence of Catalysis and pKa Values of Active Site Residues. *Biochemistry* **35**: 814-823.

Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* **41**: 207-234.

Taberman H, Andberg M, Parkkinen T, Janis J, Penttila M, Hakulinen N, Koivula A & Rouvinen J (2014) Structure and Function of a Decarboxylating *Agrobacterium tumefaciens* keto-deoxy-D-galactarate Dehydratase. *Biochemistry* **53**: 8052-8060.

Tausson WO (1928) The oxidation of waxes with microorganisms. *Biochem Z* **193**: 85-93.

Taylor AB, Czerwinski RM, Johnson WH, Jr., Whitman CP & Hackert ML (1998) Crystal structure of 4-oxalocrotonate tautomerase inactivated by 2-oxo-3-pentynoate at 2.4 A resolution: analysis and implications for the mechanism of inactivation and catalysis. *Biochemistry* **37**: 14692-14700.

Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung L-W, Read RJ & Adams PD (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D* **64**: 61-69.

Theodossis A, Walden H, Westwick EJ, Connaris H, Lamble HJ, Hough DW, Danson MJ & Taylor GL (2004) The Structural Basis for Substrate Promiscuity in 2-Keto-3-deoxygluconate Aldolase from the Entner-Doudoroff Pathway in *Sulfolobus solfataricus*. *Journal of Biological Chemistry* **279**: 43886-43892.

Thompson LC, Ladner JE, Codreanu SG, Harp J, Gilliland GL & Armstrong RN (2007) 2-Hydroxychromene-2-carboxylic acid isomerase: a kappa class glutathione transferase from *Pseudomonas putida*. *Biochemistry* **46**: 6710-6722.

Tiehm A (1994) Degradation of polycyclic aromatic-hydrocarbons in the presence of synthetic surfactants. *Applied and Environmental Microbiology* **60**: 258-263.

Vaillancourt FH, Bolin JT & Eltis LD (2006) The ins and outs of ring-cleaving dioxygenases. *Critical Reviews in Biochemistry and Molecular Biology* **41**: 241-267.

van Hylckama Vlieg JET & Janssen DB (1991) Bacterial degradation of 3-chloroacrylic acid and the characterization of cis- and trans-specific dehalogenases. *Biodegradation* **2**: 139-150.

Waddell WJ (1956) A simple ultraviolet spectrophotometric method for the determination of protein. *Journal of Laboratory and Clinical Medicine* **48**: 311-314.

Wang SC, Person MD, Johnson WH, Jr. & Whitman CP (2003) Reactions of *trans*-3-chloroacrylic acid dehalogenase with acetylene substrates: consequences of and evidence for a hydration reaction. *Biochemistry* **42**: 8762-8773.

World Health Organization (2011) Global Tuberculosis Control. *WHO Report 2011*. WHO Press, Geneva, Switzerland.

Whitman CP (2002) The 4-oxalocrotonate tautomerase family of enzymes: how nature makes new enzymes using a beta-alpha-beta structural motif. *Archives of Biochemistry and Biophysics* **402**: 1-13.

Whitman CP, Aird BA, Gillespie WR & Stolowich NJ (1991) Chemical and enzymic ketonization of 2-hydroxymuconate, a conjugated enol. *Journal of the American Chemical Society* **113**: 3154-3162.

Whitman CP, Hajipour G, Watson RJ, Johnson WH, Bembenek ME & Stolowich NJ (1992) Stereospecific ketonization of 2-hydroxymuconate by 4-oxalocrotonate tautomerase and 5-(carboxymethyl)-2-hydroxymuconate isomerase. *Journal of the American Chemical Society* **114**: 10104-10110.

Wild SR & Jones KC (1995) Polynuclear aromatic hydrocarbons in the United Kingdom environment: a preliminary source inventory and budget. *Environmental Pollution* **88**: 91-108.

Winn MD, Isupov MN & Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallographica Section D* **57**: 122-133.

Yeh HJ, Sayer JM, Liu X, Altieri AS, Byrd RA, Lakshman MK, Yagi H, Schurter EJ, Gorenstein DG & Jerina DM (1995) NMR solution structure of a nonanucleotide duplex with a dG mismatch opposite a 10S adduct derived from trans addition of a deoxyadenosine N6-amino group to (+)-(7R,8S,9S,10R)-7,8-dihydroxy-9,10-epoxy-7,8,9,10- tetrahydrobenzo[*a*]pyrene: an unusual syn glycosidic torsion angle at the modified dA. *Biochemistry* **34**: 13570-13581.

Yen KM & Gunsalus IC (1982) Plasmid gene organization: naphthalene/salicylate oxidation. *Proceedings of the National Academy of Sciences* **79**: 874-878.

Zandvoort E, Geertsema EM, Quax WJ & Poelarends GJ (2012) Enhancement of the Promiscuous Aldolase and Dehydration Activities of 4-Oxalocrotonate Tautomerase by Protein Engineering. *ChemBioChem* **13**: 1274-1277.

Zandvoort E, Geertsema EM, Baas BJ, Quax WJ & Poelarends GJ (2012) An Unexpected Promiscuous Activity of 4-Oxalocrotonate Tautomerase: The *cis-trans* Isomerisation of Nitrostyrene. *ChemBioChem* **13**: 1869-1873.